Semantics Driven Human-Machine Computation Framework for Linked Islamic Knowledge Engineering

by

Amna Basharat

(Under the Direction of Khaled Rasheed and I. Budak Arpinar)

Abstract

Formalized knowledge engineering activities including semantic annotation and linked data management tasks in specialized domains suffer from considerable knowledge acquisition bottleneck - owing to the lack of availability of experts and in-efficacy of automated approaches. Human Computation & Crowdsourcing (HC&C) methods successfully advocate leveraging the human intelligence and processing power to solve problems that are still difficult to be solved computationally. Contextualized to the domain of Islamic Knowledge, this research investigates the synergistic interplay of these HC&C methods and the semantic web and proposes a semantics driven human-machine computation framework for knowledge engineering in specialized and knowledge intensive domains. The overall objective is to augment the process of automated knowledge extraction and text mining methods using a hybrid approach for combining collective intelligence of the crowds with that of experts to facilitate activities in formalized knowledge engineering - thus overcoming the so-called knowledge acquisition bottleneck.

As part of this framework, we design and implement formal and scalable knowledge acquisition workflows through the application of semantics driven crowdsourcing methodology and its specialized derivative, called learnersourcing. We evaluate these methods and workflows for a range of knowledge engineering tasks including thematic classification, thematic disambiguation, thematic annotation and contextual interlinking for two primary Islamic texts, namely the Qur'an and the books of Prophetic narrations called the Hadith. This is done at various levels of granularity, including atomic and composite task workflows, that existing research fails to address. We leverage primarily upon students and learners engaging in typical knowledge seeking and learning scenarios. The chosen method ensures annotation reliability by introducing an 'expert sourcing' workflow tightly integrated within the system. Therefore, quantitative measures of ensuring annotation quality are woven into the very fabric of the human computation framework. The results of our evaluation demonstrate that our proposed methods are robust and are capable of generating high quality and reliable annotations, while significantly reducing the need for expert contributions.

INDEX WORDS: Qur'an, Quran, Hadith, Semantic, Islam, Islamic, Crowdsourcing, Learnersourcing, Expert-sourcing, Human Computation, Thematic Annotation, Thematic Disambiguation

Semantics Driven Human-Machine Computation Framework for Linked Islamic Knowledge Engineering

by

Amna Basharat

B.E(SE)., National University of Science and Technology, 2005

M.Sc., Brunel Univeristy, 2007

A Dissertation Submitted to the Graduate Faculty

of The University of Georgia in Partial Fulfillment

of the

Requirements for the Degree

Doctor of Philosophy

Athens, Georgia

2016

Semantics Driven Human-Machine Computation Framework for Linked Islamic Knowledge Engineering

by

Amna Basharat

Approved:

Major Professors: Khaled Rasheed
I. Budak Arpinar

Committee: Krys Kochut
Thiab Taha

Electronic Version Approved:

Suzanne Barbour
Dean of the Graduate School
The University of Georgia
December 2016

Semantics Driven Human-Machine Computation Framework for Linked Islamic Knowledge Engineering

Amna Basharat

December 2016

Inspiration

بِسْمِ ٱللَّهِ ٱلرَّحْمَٰنِ ٱلرَّحِيمِ

In the name of Allah, the most merciful, the most beneficent.

وَٱلَّذِينَ جَٰهَدُوا۟ فِينَا لَنَهْدِيَنَّهُمْ سُبُلَنَا ۚ وَإِنَّ ٱللَّهَ لَمَعَ ٱلْمُحْسِنِينَ

> And those who strive for Us - We will surely guide them to Our ways.
> And indeed, Allah is with the doers of good.
>
> — *[Al-Quran, Al-Ankaboot,29:69]*

May Allah accept this humble effort, purely for His sake, and make this a source of elevating my ranks in this world and hereafter, along with all those who have contributed in making this effort possible, foremost amongst whom are my parents, teachers, siblings, family and friends.

لِوَجْهِ ٱللَّهِ

Dedicated to

Ammi and Abbu Ji,

with all my love

Acknowledgments

ٱلْحَمْدُ لِلَّهِ رَبِّ ٱلْعَٰلَمِينَ

All thanks and praise be to Almighty Allah, the Lord of the Worlds.

سبحانك ربي ما عبدناك حق عبادتك ولا شكرناك حق
شكرك نستغفرك ونتوب اليك
سبحانك لا نحصي ثناء عليك انت كما اثنيت على نفسك
فلك الحمد والشكر

It is only due to the grace and blessings of Almighty Allah (God), that I have been blessed with the poise, the will and the perseverance to see this milestone to an end. It was only a dream when I came. Only Allah has made its realization possible.

After the prayers and the support of all my loved ones, I am deeply beholden to my advisor, Dr. Khaled Rasheed, who placed a great deal of confidence and trust

in my ideas and abilities. He gave me both independence of thought and decision making, with the perfect balance of guidance and empathy, to help me realize my dream goals that I was most passionate about. When I look back at my years spent at UGA, Dr. Rasheed stands out as someone who played the most crucial role towards my successes and growth as an individual, and towards achieving my research vision. I will always continue to derive inspiration and learning from my time spent under his fatherly supervision.

I would like to convey my immense gratitude to my co-advisor, Dr. I. Budak Arpinar, who also played a crucial role in driving my research objective and path. I am extremely grateful for the valuable suggestions, unprecedented guidance, appreciation and encouragement I received from him. He always encouraged me and challenged me to think with high level impact in mind and it certainly helped me explore avenues that I would not have been able to otherwise.

I would also like to thank my committee members Dr. Krys Kochut and Dr. Thiab Taha, for their valuable time and feedback towards refining my research objectives and direction. My years of learning at UGA will have a profound impact on my life to come.

I would like to express thanks to my colleagues Usman Nisar, Pan Huang, Kyle Krafka and Delaram Yazdansepas for their support and help. I would also like to thank the support staff at CS UGA, especially Suzi, Lafarrah, Timothy and Melody for their help and facilitation in all matters.

I would also like to express my gratitude to my Fulbright advisor, Brian Diffley, for his support always. I would like to thank the IIE and USEFP and all those who enabled this scholarship opportunity and helped me see through the completion of this degree. I would like to express my special gratitude to Ms. Rita Akhtar (Executive Director, USEFP) and those along her side, and Ms. Angela P. Aggeler (US Embassy, Pakistan), who helped me at a difficult stage in my program and ensuring everything that would help reach me a successful completion. I am extremely grateful to Dr. Aftab Maroof and Dr. Arshad Ali Shahid (FAST-NUCES, Islamabad) for facilitating, guiding and mentoring me in this time.

My very special thanks to *my Bhaiii* who took care of me all these years, played the role of 'choti ammi' and 'chota daddy', and who has always placed great pride and confidence in me and also put up with my three year old acts. This PhD would simply not have been the same without you. It makes me really proud to have you as my brother and I am grateful for all the valuable life lessons I have learnt from you bro!.

I would also like to say special thanks to my little sister, An'amta, who specially played a huge role in ensuring ease and comfort for the last few months of my PhD, when I needed all the more support and help. Her presence was indeed a source of Barakah in my time and efforts, Alhamdulillah.

A shout out to Aapi's little brothers Hebban and Zahir. Zahir, thank you for the great company, the forest walks, the great food and not to forget the endless questions and conversations. Hebban, thank you for all the wish lists and for always making sure that my luggage is loaded on my trips home.

I am extremely grateful, from the bottom of my heart, especially to so many members of my extended family, my khalas, mamus, cousins and others who always load me with duas, love and pampering throughout my life. My cousins Murrium and Umar have especially been a huge support for my travel trips throughout this period of five years.

I feel extremely fortunate and blessed to have loving friends, who may not have been physically present, yet who stood by me, supported me, cheered me on, and helped me in every possible manner they could, with their efforts, duas and foremost their love and encouragement. Mahwash, Humaira, Tahreem, Sana, Amina, Zainab Abaid, Zainab Jamil and the Baqais to name a few, I owe it to all of you. So many others who always loaded me with duas and well wishes, may Allah reward you with the very best in both worlds.

During these years, Allah also brought some special people into my life, as a special gift and mercy. Eternal friends, whom I feel so blessed to have in my life. My two pseudo moms are truly exceptional. I thank you both for bringing so much more joy, positivity and enlightenment that you bring, for all the duas, mentoring and support and for making my life and the process of my PhD in particular all the more special and cherished.

I also would like to express great admiration and love for my friend, SNG. She has been a role model, an inspiration and someone I look up to a lot and has been a mentor for me in this PhD journey.

I cannot forget all my friends here in Athens and the Atlanta community, both within and outside UGA, who made my stay comfortable and helped me in every possible way they can. I would specially like to mention Amena and Ayesha for always being there.

I also acknowledge the efforts of my ex-students. Muhammad Shoaib, was always a great source of help and support in the initial stages of my PhD. May Allah reward all of his efforts in the best manner. Also, Kamran Ali, thank you for your guidance and help at the stage I needed it most.

Bushra Abro, you may have considered yourself only a student, however, I learnt a lot from our discussions and conversations, and having you together with me as part of this journey means a lot. Thank you for sharing this passion with me, and walking together with me towards achieving the vision.

This dissertation would not have been possible without the efforts of many nameless 'crowds' and 'learners' and 'experts' who participated in the tasks and made this research an astounding success. In particular the students at Alhuda International, Alhuda E-Campus and those part of the Alhuda Apps group played a significant role in making this study truly worthwhile.

In particular, I would like to thank Dr. Farhat Hashmi (CEO, Alhuda International), Dr. Kanwal Kaisser (Director, HB Alhuda Int.), Sr. Nabila Salimi (Alhuda International), Sr. Rifqah Asghar *(Alhuda International)*, Ms. Ayesha Basheer *(Lecturer, IIU and Islamic Online University)*, Ms. Asma Batool Naqvi (HOD DCS&SE, FBAS, IIUI), Dr. Shermeen Nizami and Sr. Doha Hijab *(Alhuda E-Campus)* and Sr. Shazia Nawaz *(Alhuda US)*, for facilitating the research study. I am also thankful to the authors of sunnah.com for providing us with the valuable datasource to carry out this research.

I would also like to thank Dr. Adel Amer, for his time, feedback, guidance and expert contributions. Many thanks to Matib Ahmed, Khalifa Al-Jeddah, Omar

Iqbal and others, who are too many to be named, for their contributions.

Special thanks to Sr. Taimiyyah Zubair, Br. Imran Haq and Sr. Mariam Wasti *(Alhuda Canda)* for meeting with me during my trip to Canada. I would also like to thank Hisham Al-Hadi for his time to discuss my research, endorsing it, facilitating the study forward and the positive encouragement.

There are others, who prayed and wished well, without knowing even me knowing. May Allah reward them all with the best and highest rewards in both worlds. Ameen.

I must say this PhD journey has been truly unique. It started with a passion, desire to do work for a purpose, unprecedented and unparalleled. It was made possible by many but foremost the following:

My dearest and most beloved parents, for whom, no words of love and gratitude can ever be enough. They instilled in me the drive, the determination and the vision to always excel, seek higher and give my best at everything I do. They have truly empowered me with the best, in every way. They place more pride in me and my abilities than my own self. It was my father's dream to see me complete a PhD from the US. Allah blessed me by enabling me to fulfill my dream through his dream. There are no words of gratitude that can suffice for all the unconditional love, care, support, duas and selfless sacrifices my parents have made for me and my siblings through out their lives. Their hard work and poise towards providing us with the very best has been unprecedented. May Allah enable me to become the best Sadqa-e-Jariah for them and may Allah bless them infinitely in this world and the next with best rewards and elevate them to highest ranks.

My PhD is special because Allah enabled me to work on something that I was truly passionate about. Allah SWT greatly blessed me by enlightening my heart and life with the Qur'an. The foremost source of this inspiration is my dearest and most beloved teacher, Dr. Farhat Hashmi, whom Allah made the source of bringing this enlightenment and guidance in my life. I cannot thank her enough for all her love, duas and support and for placing immense trust and faith in me and facilitating me to the maximum possible extent. May Allah honor her and reward her in the best manner in both worlds, and all those who are alongside her, and especially those who encouraged, supported, facilitated and participated in my work. May Allah grant her infinite blessings and rewards for all her support, mentoring and role modeling.

Last but not the least, I had always believed that Allah appoints special angels to look after us. This dissertation would be incomplete without the mention of one special angel who was sent my way during the span of this PhD, through whom, I truly believe the experience of not only this PhD, rather my entire life has been transformed. Words would not do justice to describe this transformational & enriching experience. I cannot thank you enough for standing by me, for believing in me, for having so much faith & confidence in me, for virtually living and traveling with me, & for owning me *(& also my thesis)* like your own. Words are neither enough to express gratitude enough nor express the bliss and happiness experienced from the jewels of love, care, support, duas, appreciation & motivation you have showered upon me. May Allah preserve this relation eternally *(for His sake alone)* & guide our way to the highest levels in Jannah together. Ameen.

لله ... في الله ... بالله

Contents

Acknowledgments **viii**

List of Figures **xx**

List of Tables **xxiv**

1 Introduction **1**

1.1 Research Context and Motivation 2

1.2 Research Novelty and Contributions 5

1.3 Dissertation Organization 7

2 A Vision and Conceptual Framework for Linked Islamic Knowledge **9**

2.1 The Vision of Linked Islamic Knowledge 10

2.2 The LOD Potential for Islamic Knowledge 13

2.3 A Generic, High-Level Conceptual Framework For Linked Open Islamic Knowledge 19

2.4 The Need and the Role for Human Computation and Crowdsourcing 24

2.5 Conclusions 26

3 Human Computation and Crowdsourcing meet the Semantic Web: A Survey **28**

3.1 Introduction 30

3.2 Background 34

3.3 Application of Human Computation Techniques to Semantic Web Research 49

3.4 A Collective Intelligence Genome for the Semantic Web 69

3.5 Crowdsourcing based Workflow Systems for the Semantic Web . . . 76

3.6 Application of Semantic Web Techniques to Facilitate Human Computation 90

3.7 Discussion and Analysis 91

3.8 Conclusions 101

4 Text Mining Verse Similarity for Multi-lingual Representations of the Qur'an **104**

4.1 Introduction 105

4.2 Verse-to-Verse Similarity Computation Framework for the Qur'an . 108

4.3 Experimentation and Results 115

4.4 Discussion and Analysis 124

4.5 Conclusion 125

5 Crowdsourcing Framework for Linked Data Management **126**

5.1 Introduction 127

5.2 Crowd-based Task and Workflow Management Architecture for the LOD 134
5.3 Experimental Evaluation 146
5.4 Comparison with Related Works 160
5.5 Beyond CrowdLink: Using CrowdSourcing in the LOD 163
5.6 Conclusions 165

6 A Semantic Framework for Thematic Annotations of the Qur'an Leveraging Crowds and Experts 166

6.1 Introduction 168
6.2 Problem Definition: Thematic Annotation in the Qur'an 171
6.3 Hybrid Architecture for Harnessing Crowd and Expert Annotations 174
6.4 Experimental Evaluation 186
6.5 Conclusions 194

7 Learnersourcing Thematic and Inter-Contextual Annotations from Islamic Texts 196

7.1 Introduction 197
7.2 Background and Related Work 201
7.3 Research Approach: Semantics Driven Learnersourcing Methodology 206
7.4 Dynamic Learnersourcing Workflow Design for Thematic Linking and Annotation of Hadith 208
7.5 Experimentation and Evaluation 222

7.6 Discussion and Conclusions 229

8 Semantic Hadith: A Framework for Publishing Linked Hadith Data 231

8.1 Introduction 232

8.2 Ontology for Semantic Hadith 238

8.3 Link Modeling and Design Issues 241

8.4 Linked Data Publishing Framework and Implementation 247

8.5 Prospective Applications 254

8.6 Related Work 257

8.7 Conclusions and Future Work 258

9 Conclusion 260

9.1 Research Summary and Contributions 261

9.2 Research Significance and Potential Impact 265

9.3 Future Research 266

Bibliography 269

List of Figures

1.1 Challenges in Knowledge Engineering Processes 3

2.1 The Macro Structure of Linked Islamic Knowledge Landscape . . . 15

2.2 High level conceptual framework for Linked Islamic Knowledge Generation, Publishing and Consumption 20

3.1 Ontology Development Process (simplified adaptation from [153]) . 36

3.2 Classification of Tasks in the Ontology Engineering Process (as summarized in [186]) 44

3.3 Classification of Tasks in the Process of Ontology Development (as summarized in [186]) 45

3.4 The Typical Task Lifecycle in a Crowdsourcing Platform 52

3.5 Classification of Human Computation Tasks and Genres in Semantic Web (Ontology Engineering) 60

3.6 Classification of Human Computation Tasks and Genres in Semantic Web (Linked Data Management) 63

3.7 Main Stages in using the uComp Plugin from [5] 68

4.1 Qur'anic Verse Similarity Computation framework 108
4.2 Results for all representations (Similarity Range vs. No of Verse Pairs) 123

5.1 High-level Architecture of the Semantics-Enriched Workflow and Task Management Engine 135
5.2 The Process of Task Specification 136
5.3 The Process of Task Generation and Publishing 137
5.4 The Process of Task Review and Results Retrieval 143
5.5 The Process of Updating the Ontology from the Processed Results . 143
5.6 A Human Intelligent Task (HIT) in AMT 148
5.7 Evaluation Results for Eleven Task Groups 151
5.8 Confidence based Analysis for Entity Disambiguation Tasks (Grouped by TaskID) 156
5.9 Confidence based analysis for Entity Disambiguation tasks (Grouped by WorkerID) 157

6.1 A segment of the thematic annotation knowledge model populated via the crowd 173
6.2 Hybrid Architecture for Harnessing Crowd and Expert Annotations 174
6.3 An Example of Learner tasks introduced in the learning scenarios (Thematic Disambiguation) 176
6.4 Human-Machine Computation Workflow for Disambiguation Task . 177
6.5 An Example of Learner task (Thematic Annotation) 178

6.6 Human-Machine Computation Workflow for Annotation Task . . . 179
6.7 Task Design for Thematic Annotation of Qur'anic verses 181
6.8 Decision Model used for Approving the Tasks based on Crowd Submissions and marking the ones to be sent to the Experts as Reviewable 182
6.9 Distribution of Number and percentage of Disambiguation Task from Crowd and Expert Sourcing stage 187
6.10 Distribution of Number and percentage of Annotation Task from Crowd and Expert Sourcing stage 188

7.1 A typical path of knowledge synthesis followed by Qur'anic learners 199
7.2 Overview of the Learnersourcing Methodology 206
7.3 Workflow Design 210
7.4 A snapshot of the look up hadith task, divided in three dynamic steps 215
7.5 A snapshot of the annotate Hadith task, divided in two dynamic steps 218
7.6 Decision Metrics for Task 1 220
7.7 Task Lifecycle 222
7.8 Number of tasks that reached agreement for different number of submissions 225
7.9 Performance patterns for two separate learners for Task 1 228

8.1 A snapshot of a typical Qur'anic Commentary 233
8.2 A Sample Hadith Snapshot 236
8.3 High Level Design of Semantic Hadith Ontology 239
8.4 Conceptual Design model for Hadith-Hadith Relationship 242

8.5 Conceptual Design model for Hadith-Verse Relationship 244
8.6 Conceptual Design model for Verse to Hadith Relationship based on Scholarly Commentaries 246
8.7 A view of the proposed and available Linked Data Cloud for Islamic Knowledge Sources 249
8.8 Linked Data Generation and Publishing Framework for Semantic Hadith 253

List of Tables

2.1 Summary of Related Works towards Islamic LOD 19

3.1 Broad Classification of Human Computation Tasks in the Semantic Web (Ontology Engineering) and the Respective Genres 56

3.2 Broad Classification of Human Computation Tasks in Semantic Web (Linked Data Management) and Respective Genres 62

3.3 Overview of the Collective Intelligence Genome 71

3.6 Comparative Analysis of Crowdsourcing Research in the Semantic web 82

3.4 Application of Collective Intelligence Genome to Semantic Web Research 102

3.5 Threads and Dimensions of Analysis 103

4.1 Dataset representations and samples 109

4.2 Similarity Classes for Analysis 114

4.3 Precision and Recall Measures 116

4.4 Analysis of false positive case (Q-ENG) 120

5.1 A Sample Task Specification (Review Type) 138
5.2 Sample Task Data Input Generated from the Query Processor . . . 140
5.3 Sample Task Assignment Properties 141
5.4 Eleven Task Groups Used in the Experimental Evaluation (G: Group ; P: Participants ; R: Responses) 150
5.5 Attributes and Answers in a Result file for GetStateNameforRiver Task 153
5.6 Candidate Entities from ACM and DBLP for Entity Disambiguation. 157

6.1 Notation used in the Decision Analytics Process 180
6.2 Agreement and Confidence Thresholds for Crowd and Expert Output Decision Analytics 185
6.3 Results for Disambiguation and Annotation Tasks for both stages . 186
6.4 Crowd Submission Scoring for Disambiguation Task 191
6.5 Crowd Submission Scoring for Annotation Task 192
6.6 Results showing the number of submissions as evaluated based on the system and the expert responses 193
6.7 Comparative Analysis of Crowd and Expert Tasks (Annotation Tasks) 193

7.1 Results for All Tasks 223
7.2 Learner Submission Classification for all tasks 226

8.1 Entity Statistics in the Semantic Hadith Dataset(Sunnah.com) . . . 252
8.2 Link Statistics in the Semantic Hadith Dataset 252

Chapter 1

Introduction

إِنَّا هَدَيْنَٰهُ ٱلسَّبِيلَ إِمَّا شَاكِرًا وَإِمَّا كَفُورًا

Indeed, We guided him to the way, be he grateful or be he ungrateful.

— *[Al-Quran, Al-Insaan,76:3]*

Chapter Overview[1]

This chapter provides an overview of the dissertation by laying out the research context and describes the motivation behind this research. The research novelty and contributions are highlighted and a dissertation roadmap is provided to facilitate the reader.

[1]Partially appears as:
Amna Basharat. "Semantics Driven Human-Machine Computation Framework for Linked Islamic Knowledge Engineering", pages 793-802. Springer International Publishing, 2016.

1.1 Research Context and Motivation

One of the major challenges hindering successful application of ontology-based approaches to data organization and integration in specialized domains is the so-called *'knowledge acquisition bottleneck'*[162]- that is, the large amount of time and money needed to develop and maintain the formal ontologies. This also includes ontology population and semantic annotation using well established vocabularies. This research primarily is motivated to overcome the inherent knowledge acquisition bottleneck in creating semantic content in semantic applications. We have established how this is particularly true for knowledge intensive domains such as the the domain of Islamic Knowledge, which has failed to cache upon the promised potential of the semantic web and the linked data technology; standardized web-scale integration of the available knowledge resources is currently not facilitated at a large scale [29]. To date, only one dataset on the Linked Open Data (LOD) cloud in the domain exists [168].

To understand the knowledge acquisition bottleneck encountered in this domain (and others), consider the knowledge engineering processes illustrated in Figure 1.1. Existing methods towards semantic annotation and linked knowledge generation are either a) computationally driven, employing on text mining and information extraction methods or, 2) expert driven (such as conceptual modelling, annotation and validation). While the computational methods may assist in large-scale knowledge acquisition, however, the lack of formalized and agreed upon knowledge models and sensitivity of the knowledge at hand- primarily obtained from unstructured and multilingual data- makes the knowledge engineering

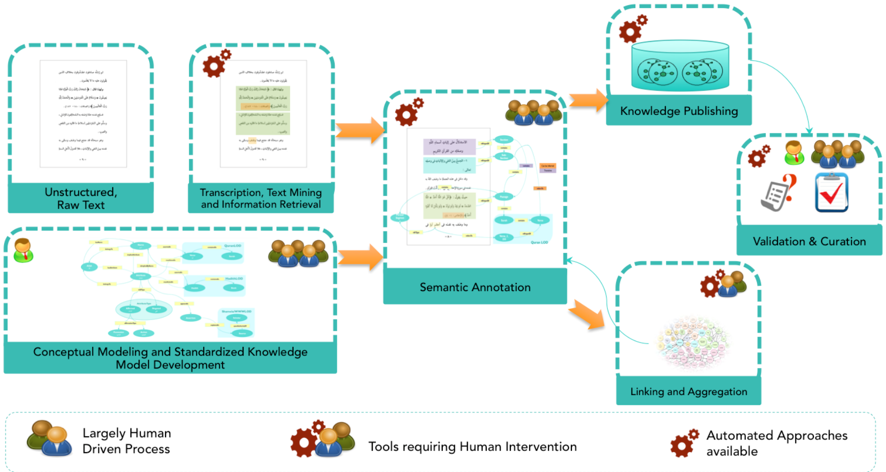


Figure 1.1: Challenges in Knowledge Engineering Processes

process far from trivial. Islamic domain suffers a great deal from a lack of suitable training data and gold standards. This only adds to the challenge of ensuring the reliability and scalability of these methods. Expert driven methods involve subject specialists however, these are often not scalable (time or cost). It is no wonder that the efforts towards the vision of standardization and formalization of Islamic Knowledge as proposed by [18] have remained futile.

To address these and similar challenges, researchers have recognized that the realization of the semantic and linked data technologies will require not only computation but also significant human contribution [186],[182]. Humans are simply considered indispensable [54] for the semantic web to realize its full potential. Emerging research is advocating the use of Human Computation & Crowdsourcing (HC&C) methods to leverage human processing power to harness the 'collective intelligence' and the 'wisdom of the crowds' by engaging large number of online contributors to accomplish tasks that cannot yet be automated. There has been growing interest to use crowdsourcing methods to support the semantic web and linked open data research by providing means to efficiently create research relevant data and potentially solve the bottleneck of knowledge experts and annotators needed for the large-scale deployment of semantic web and linked data technologies. Recent framework called CrowdTruth proposed by [100] is a step forward that recognizes the challenges in gathering distributed data in crowdsourcing.

Within the realm of (semantic web based) knowledge engineering tasks, several levels of complexity may be encountered. Some tasks are simple, while others are more knowledge intensive. While some may reasonably be amenable to computational approaches, others need domain specific expert annotations and judgements. Therefore, not all tasks are fit for general purpose crowdsourcing. This is specifically true for the domain of islamic knowledge, owing to the specialized nature of the learning needs presented by diverse users, and heterogenous, multilingual knowledge resources. Therefore, we emphasize that specialized domains such as the one that forms the basis of this research, i.e. the Islamic knowledge domain, needs more than just faceless crowdsourcing. Emerging research paradigm recognizes this in the form of nichesourcing [49], [149] or Expert-Sourcing [158] , as a natural step in the evolution of the crowdsourcing to address the need of solving complex and knowledge intensive tasks, and as means to improve the quality of crowdsourced input.

1.2 Research Novelty and Contributions

We believe this to be the first of its kind research that attempts to formally model, interlink and synthesize historical and multi-lingual religious texts using novel human computation methods and workflows. Through this research, we make the following primary contributions:

With respect to the domain of Semantic Web and Human Computation, we successfully leverage the human computation methods including crowdsourcing and learnersourcing to overcome the knowledge acquisition bottleneck in knowledge engineering tasks.

- We design and implement a novel framework that inter-leaves human and machine intelligence, driven by ontologies, to facilitate knowledge acquisition, semantic annotation and linked data management tasks. Our framework utilizes the notion of semantics based task profiles and workflows, to augment computational approaches of knowledge extraction, with crowd contributions. Our crowdsourcing framework is fully generic and tested for various linked data management tasks such as link verification, entity disambiguation and knowledge acquisition (Chapter 5).

- We developed a semantics based task specification model allows for composing automated knowledge acquisition workflows that combine machine computation with crowdsourced annotations at two levels. At the lower level, simple and less knowledge intensive tasks are crowdsourced using the Amazon Mechanical Turk platform. At the higher level for more knowledge intensive tasks, skilled workers and experts are engaged through a custom web portal, designed with a citizen science approach. We demonstrate the effectiveness of this hybrid model for several key knowledge engineering tasks such as thematic disambiguation and thematic annotation, in the Qur'an. Results of our study demonstrate that we can obtain high quality annotations reliably through our proposed technique (Chapter 6).

- We further enhanced our crowdsourcing model, and implemented a specialized form of crowdsourcing called learnersourcing to allow for tasks to be adaptively delegated to learners or students engaged in learning scenarios pertinent to the domain. The learnersourcing methodology is validated by implementing a composite task based knowledge acquisition workflow. The results of the study show promising results by providing highly reliable and accurate annotations for thematic and inter-contextual interlinking of Hadith data. The reliability of the method is ensured by tightly integrating expert sourcing workflows within the system. A key highlight of this model is the ability to reduce the amount of tasks requiring experts through the application of adaptive task delegation thresholds to the learners (Chapter 7).

With respect to the domain of Islamic Knowledge, we undertook case studies in order to validate the design and implementation of the framework, resulting in several contributions.

- We propose a vision, a methodology and a conceptual framework for addressing knowledge formalization, acquisition and interlinking challenges in the Islamic domain (Chapter 2).

- We designed and implemented a similarity computation framework for interlinking entities such as verses in the Qur'an or hadith. This text mining framework based on similarity computation serves as the foundation for providing candidate seed input to the crowdsourcing and learnersourcing workflows (Chapter 4).

- We design and implement knowledge acquisition, interlinking and publishing workflows for this domain, which are entirely reusable and customizable and are expected to be used in the future for obtaining semantic annotations from a range of knowledge sources (Chapter 6 and 7).

- We propose knowledge models and vocabularies for formally modeling and interlinking the two primary sources of the Islamic domain, namely the Qur'an and the Hadith, which form the basis of all other texts. We also expand the linked data cloud by providing a new datasource called 'Semantic Hadith' and creating richer knowledge links with available data sources (Chapter 8).

1.3 Dissertation Organization

The rest of the dissertation is divided into the following chapters:

Chapter 2: A Vision and Conceptual Framework for Linked Open Islamic Knowledge

Chapter 3: Human Computation and Crowdsourcing meet the Semantic Web: A Survey

Chapter 4: Text Mining Verse Similarity for Multi-lingual Representations of the Qur'an

Chapter 5: Crowdsourcing Framework for Linked Data Management

Chapter 6: A Semantic Framework for Thematic Annotations of the Qur'an Leveraging Crowds and Experts

Chapter 7: Learnersourcing Thematic and Inter-Contextual Annotations from Islamic Texts

Chapter 8: Semantic Hadith: A Framework for Publishing Linked Hadith Data

Chapter 9: Conclusions

Chapter 2

A Vision and Conceptual Framework for Linked Islamic Knowledge

إِنَّ ٱلَّذِينَ قَالُوا۟ رَبُّنَا ٱللَّهُ ثُمَّ ٱسْتَقَـٰمُوا۟ تَتَنَزَّلُ عَلَيْهِمُ ٱلْمَلَـٰٓئِكَةُ
أَلَّا تَخَافُوا۟ وَلَا تَحْزَنُوا۟ وَأَبْشِرُوا۟ بِٱلْجَنَّةِ ٱلَّتِى كُنتُمْ تُوعَدُونَ ٣٠

Indeed, those who have said, "Our Lord is Allah " and then remained on a right course - the angels will descend upon them, [saying], "Do not fear and do not grieve but receive good tidings of Paradise, which you were promised.

— *[Al-Quran, Fussilat,41:30]*

Chapter Overview[1]

This chapter presents the emerging vision and the need for Linked Open Islamic Knowledge, its requirements and foundations. The important challenges and limitations of the existing research landscape are highlighted in an attempt to analyze why it fails to meet the requirements of meeting the vision at hand. A conceptual framework that would functionally fulfill the Islamic LOD vision is presented. A brief insight is also provided to hint at the need and the role of Human Computation & Crowdsourcing (HC&C) methods and how they may facilitate the realization of the proposed vision.

2.1 The Vision of Linked Islamic Knowledge

The emerging technologies in the recent years have greatly revolutionized the different ways to interact with knowledge. There is an increasing need to search for new ways of modeling, standardizing, aggregating, linking, publishing, visualizing and presenting knowledge for Islamic and religious knowledge providers and seekers, to engage, facilitate and educate them. So far web-scale integration of Islamic knowledge resources is not facilitated, mainly due to the lack of adoption of shared principles, datasets and schemas. This research, therefore, aims to investigate how Linked Open Data (LOD) technologies can solve the problem of information integration and provide new ways of teaching and learning Islamic knowledge. The

[1]This chapter is published as:
Basharat, A., Rasheed, K., Arpinar, I.B. "A Conceptual Framework for Linked Open Islamic Knowledge", *In: International Journal on Islamic Applications in Computer Science And Technology âĂŞ IJASAT (2016).* Vol 4, No 2. 16-25.

linked data approach has emerged as the de facto standard for sharing the data on the web. The term âĂIJlinked dataâĂİ refers to a set of best practices for publishing and connecting structured data on the web [37]. The linked data design issues provide guidelines on how to use standardized web technologies to set data-level links between data from different sources[89]. Increased interest in the LOD has been seen in various sectors e.g. Education [53], [155], Scientific research [16], libraries [82], [95], Government [56], [90], [166], Cultural heritage [63] and many others, however, the religious sector has yet to cache upon the power of the linked open data.

2.1.1 Motivation for an LOD Approach for Islamic Knowledge

Research in computational informatics applied to the Islamic knowledge has primarily centered around Morphological annotation of the Qur'an [60], [59], Ontology modeling of the Qur'an [7], [21], [65], [211], [212], and Arabic Natural language processing [65]. The LOD take-up in the area of Islamic knowledge has been particularly extremely limited. To date only one recent study has attempted to publish a small data-set pertaining to Qur'anic knowledge [168]. Thus, LOD opens the opportunity to address information integration and interoperability issues in the field. According to the assessment of this research, the suitability of LOD paradigm to the domain of Islamic knowledge stems from several groundings foremost of which includes a great need for standardization, common knowledge models and vocabularies in the domain of religious learning in general and Islamic knowledge in particular. In addition, there is not only great potential for generating 'links' in the distributed content that has been made available over the years with the advent of technology and the internet, there is also dire need to establish these links to better facilitate relevant and timely knowledge discovery, aggregation and efficient retrieval by knowledge seekers and educators alike. The LOD vision promotes open and interoperable access to knowledge, along with the support for multi-lingual content, both of which are highly suitable to the needs of the Islamic knowledge.

2.1.2 Contributions of the Chapter

This chapter makes the following contributions:

- It provides insights into understanding the vision, the context, the need and potential for Linked Open Islamic Knowledge; it explains the proposed Macro-structure of linked Islamic knowledge and provides the classification and the nature of links at the knowledge level (Section 2.2).

- It provides an understanding of the requirements for achieving this vision, and the required foundations by proposing a high-level conceptual framework that would need to be developed to achieve this vision; it also delineates upon the key challenges that are encountered in the realization of such a framework (Section 2.2.2).

- It examines the role of Human Computation and how it may be used in overcoming the key challenges associated with resolving the knowledge acquisition bottleneck in order to achieve the LOD vision for Islamic knowledge (Section 2.4).

2.2 The LOD Potential for Islamic Knowledge

The major challenge for the Islamic knowledge domain is to start adopting LOD principles and vocabularies while leveraging on existing data available on the web. However, in order to understand how and if at all the LOD paradigm is suitable to this domain, it is important to understand the structure of Islamic knowledge, some unique requirements for the knowledge providers and seekers in the domain and the nature of potential links at the knowledge level.

2.2.1 Understanding the Macro Structure of Islamic LOD

Figure 2.1 shows a broad overview of the macro structure of the Islamic knowledge landscape that has evolved since the revelation of the Qur'an. As shown in Figure 2.1, the Qur'an is the primary book of knowledge at the center of all Islamic knowledge. It forms the foundation for anyone wanting to learn Islam be it a Muslim or Non-Muslim. The second most important source is the books of Hadith (Narrations of the Prophet Muhammad), which are also considered as the primary source of knowledge. Together, these two are the most important sources of knowledge. The next level are the scholarly books of Qur'anic commentaries and Exegesis, which contain interpretations and explanations to the verses in the Qur'an based on scholarly interpretation. Scholars over the years have written and compiled thousands of books on Qur'anic Exegesis. Some are more prominent than the others based on the merit of scholarly descent, caliber and the authenticity of the material cited. These books heavily rely on the books of Hadiths and the Qur'an itself for explaining different aspects of the verses. At the next level are classical books, written in the early periods, which rely heavily on these commentaries, the books of Hadith and the Qur'an. A detailed classification and discussion of Qur'anic explanation is provided by Philips [154]. Although, new research, literature and educational content has continued to be generated with a high pace, and over the recent years much of it has been made available in the digital format, however it remains firmly rooted in its links to primary, secondary and tertiary sources of knowledge. Over the recent years, several structured data repositories and multimedia content in the form of audio and video lectures from contemporary scholars have also been made available which contain references to these classical and primary sources of reference. Therefore, the content is rich in 'potential links', which existing repositories and applications fail to capture.

Existing Islamic knowledge sources and applications are little known to be interoperable, and do not encourage automated discovery, recommendation and synthesis. As a result, teachers, students and self-learners spend too much time looking for resources or they spend too little and may end up making decisions based on incomplete information. The user experience of a knowledge seeker or an educator in Islam can be greatly improved if the system runs on a backend supported with linked open data for the representation of Islamic knowledge. In this research an effort is made to examine how the LOD can be adopted to model

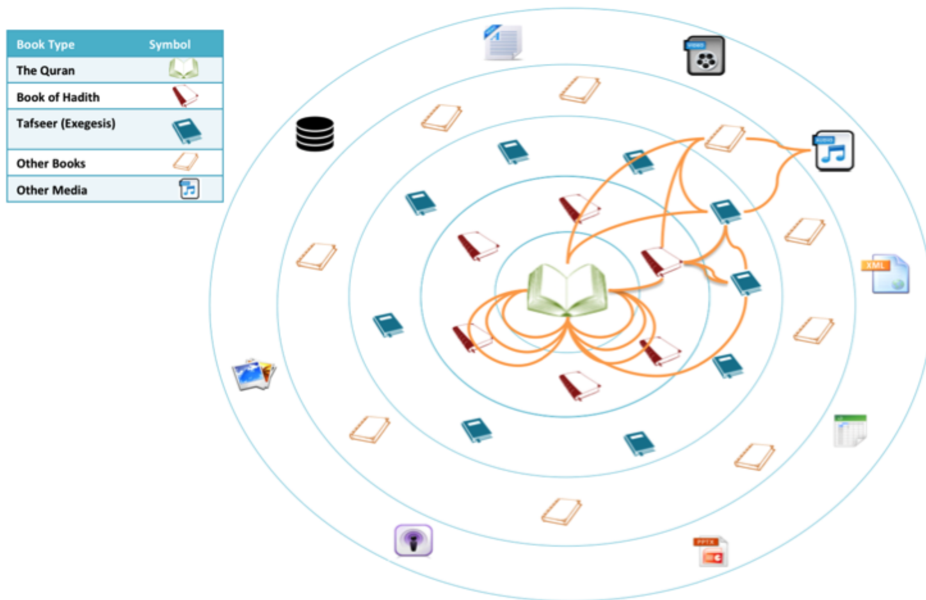


Figure 2.1: The Macro Structure of Linked Islamic Knowledge Landscape

Islamic knowledge to address some major challenges and needs that arise for students, teachers, researchers and scholars of this domain.

2.2.2 The Need and Challenge for Linking Islamic Knowledge

It is important to understand why linking of relevant knowledge sources is important and particularly challenging in the domain of Islamic knowledge. The Qur'an is the primary book of knowledge. It is not possible to understand the Qur'an in isolation. Qur'anic understanding is incomplete without referring to the Hadith and books of exegesis. Cross-linking of resources will facilitate learners to cross-navigate knowledge sources at various levels of granularity to ensure a more efficient learning experience. Similarly, contemporary literature must be backed by credible primary sources; however, there is no standardized means by which this cross-linking may be made possible. Often when referring to a particular opinion, chain of narrators and scholars is important and must be traceable in order to ensure the authenticity of knowledge being quoted, however, often times this is not accurately cited in books and becomes a challenge for the learners to verify. Another level of complexity is the difference is schools of thought that make the knowledge selection and reliability difficult. While the religious practices need to be backed by primary sources, there is little or no support for cross comparison because knowledge sources are neither synthesized nor linked in such a manner. These are just some of the unique requirements for a knowledge seeker or an educator of knowledge in Islam and therefore imply the strong need for linking the relevant knowledge sources, thus making their access, retrieval and analysis easier. However, the most challenging aspect that emerges as a result is that fulfilling these needs through purely computational approaches is not an easy task. Human intelligence plays an indispensable role in capturing and modeling the knowledge in the right manner.

2.2.3 The Nature of Potential Links

In Figure 2.1, where the Macro knowledge structure is shown for the Islamic domain, the potential links may be created at various levels, depending on the level of granularity chosen for modeling the knowledge units. This research proposes the classification of these knowledge units and their respective links at two levels: Macro and Micro Level.

Macro-Level Links

In the proposed Islamic LOD based knowledge framework, Macro-Level Units are considered at the level of granularity that is equivalent to a Verse in the primary source i.e. the Qur'an. There may be verses in the Qur'an that link to other verses. There is also evidence in the scholarly interpretations of inter-chapter relationships. Different verses help to explain one another. In fact, the primary principle of Qur'anic Exegesis states that the Qur'an itself does the best explanation of the Qur'an [154]. Therefore, most crucial links would tend to be the Verse-to-Verse links. Scholars over the centuries have attempted to extract these links and are captured in their commentaries. On a similar note, the next most important level of links to be captured is Verse-to-Hadith Relations. Qur'anic interpretation heavily relies on the narrations of the Prophet Muhammad, called Hadith. Modeling these formally using computational knowledge models is not a trivial task, however, if done, this would greatly facilitate Qur'anic scholarly pursuits. It is also important to consider a Passage (Groups of Verses) as an important macro level entity, when modeling relationships between different entities at varying levels of granularity. Qur'anic verses often cannot be interpreted in isolation, and need to be closely studied in combination with the context and the surrounding verses they appear with. On a similar note, other higher-level constructs such as passages, sections and sub-sections that belong to some text may become part of some relation that is essential to capturing the link. To date, to the best of the background review that this research has undertaken, no formal knowledge model is known to have captured links and relations at this level.

Micro-Level Links

Since the language of the Qur'an and Hadith is Arabic, much of the classical literature is also in the Arabic language. Arabic is a rich language, with complex morphology, and semantics [65]. There is no bound to the word-to-word relations that may exist in these texts. Different words have common roots, which appear differently with different meanings in different contexts (polysemy). Similarly words with related meanings may not have common roots. There may be words linked due to their morphology. Related words may appear in the same verse, same surah (chapter) or across chapters. Capturing links at this level would constitute, what has been coined in this research as, Micro-level links. In addition to words and roots, often times, portions of large verses are used in literature and quoted as a significant phrase or a segment. These would also be classified as micro-level links.

2.2.4 Related Work towards Islamic LOD

There have been some recent works, which have attempted to create a foundation for the Islamic LOD. Amongst these, noteworthy mentions are SemanticQuran [169] and QuranOntology [78]. The recent work of the authors of this research demonstrates the modeling and implementation of macro-level links namely, Verse-

Table 2.1: Summary of Related Works towards Islamic LOD

Research Work	Types of Links in the Dataset	Original Data Source(s)
Semantic Quran [169]	**Micro-Level Links** Entity Links with DBPedia	Quran Corpus (quran.com) Tanzil.net
Quran Ontology [78]	**Macro-Level Links** Verse-Verse Relations Verse-Topic Relations	Quran Corpus (quran.com) Qurany [2] Semantic Quran [169] QurSim [167]
Semantic Hadith [30]	**Macro-Level Links** Verse-Hadith Relations Hadith-Verse Relations Hadith-Hadith Relations	Sunnah.com QuranOntology [78] Semantic Quran [169]

Hadith links in the work entitled Semantic Hadith [30] . A summary of these works, the types of links they capture and the data sources they have relied on are given in Table 2.1.

2.3 A Generic, High-Level Conceptual Framework For Linked Open Islamic Knowledge

As related research in other domains suggest, linked data technologies can help to integrate the work of disperse institutions producing diverse linked data [155]. Over the years, a number of open applications and digital repositories have been made available containing a wealth of classic Islamic literature including the Qur'an,

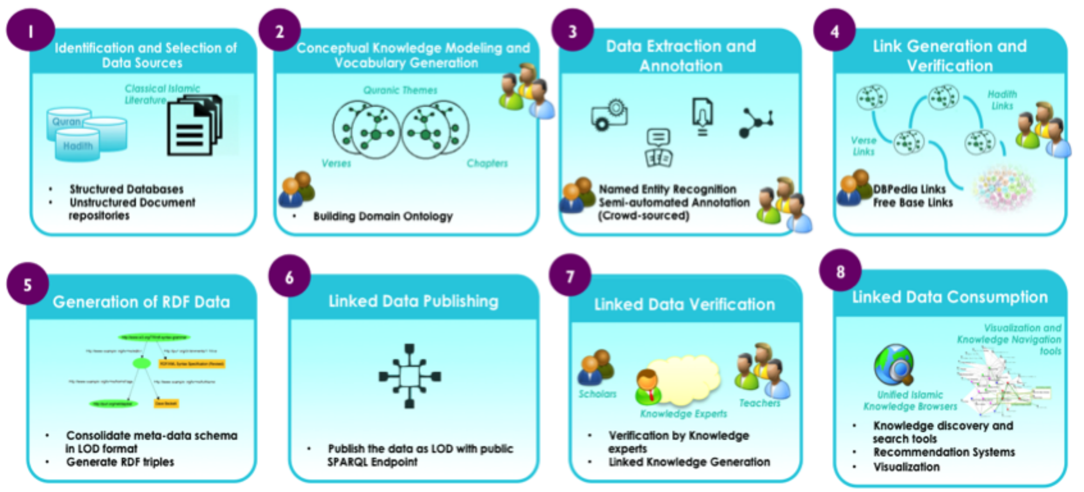


Figure 2.2: High level conceptual framework for Linked Islamic Knowledge Generation, Publishing and Consumption

Books of Hadith, Fiqh, Tafseer, Aqeedah and many others. Realizing the vision of linked Islamic knowledge requires primarily a linked data approach on Islamic knowledge repositories and establishing a framework for bringing this knowledge into a more interoperable, integrated one for sharing, connecting and discovering knowledge. Linked Open datasets need to be created by harvesting information from isolated datasets. By using semantically interoperable linked data, these materials and repositories could be cross-connected to other repositories and portals, thereby allowing a variety of information systems to disseminate knowledge across platforms. This would be crucial in facilitating students, teachers, researchers and general knowledge seekers of Islam in particular and religion in general. Figure 2.2 shows the high-level conceptual framework that would need to be developed in order to achieve this vision.

Functional Stages of an Islamic LOD Framework

The functional aspects of this framework are briefly outlined.

- ***Identification and Selection of Data Sources***: This stage would involve the identification and selection of appropriate, heterogeneous data sources to determine the scope of the content. Data sources may be structured or unstructured. Any existing semantic models or potential reusable entities may be identified.

- ***Conceptual Knowledge modeling and Vocabulary Generation:*** To accomplish this, vocabularies in the form of ontologies about the core knowledge content would need to be devised, or reused if available, as well as metadata schemas for knowledge representation appropriate for the data sources. A key research challenge will be accomplishing the creation of semantically unambiguous metadata as automatically and accurately as possible.

- ***Data Extraction and Annotation:*** This would involve entities within the selected texts to be annotated and further information about these entities may be retrieved from links from existing repositories e.g. applications built using the proposed system in this research could automatically suggest further information about the entity from DBPedia or Freebase and related educational materials from other institutes.

- ***Link Generation and Verification:*** By semantic data aggregation, knowledge based links will be generated e.g. Qur'anic verses would be potentially linked with other verses, related Hadith and extracts in the books of tafseer that will enable better knowledge navigation and understanding while undertaking serious study of the Qur'an.

- ***Generation of RDF Data:*** This stage would be one of the key stages prior to publishing the data as linked data and would involve conversion and generation of RDF compatible data. A linked data set will convert existing knowledge into RDF Triples using pre-defined vocabularies and this can be done for Islamic knowledge at various levels. When done across various datasets, matching across RDF triples will result in linking that will become the means to offer a richer knowledge experience.

- ***Publishing Linked Data:*** Linked data will be published over the LOD cloud, and linked across the existing data sets. The data will be discoverable via SPARQL query endpoints.

- ***Link Data Verification:*** A crucial requirement is the verification accuracy of the published data and ensuring the authenticity and soundness of not only the knowledge itself but also the links generated. For this purpose complete reliance on automated means is not possible given the sensitivity of Islamic knowledge. Human intelligence will need to be relied upon. Means to achieve this would need to be devised such as some specialized crowd-sourcing experiments or collaborative annotation and verification.

- ***Linked Data Consumption:*** The potential utilization of the generated linked data will need to be shown via some applications and visualizations.

This would be crucial towards validating the success of the proposed vision of Islamic LOD.

Existing Challenges

The domain of Islamic knowledge, has failed to cache upon the promised potential of the semantic web and the linked data technology. Established research has recognized that ontology-based approaches to data organization and integration in specialized domains have produced significant successes [147]. One of the major challenges hindering such approaches from successful application to the Web at large is the so-called 'knowledge acquisition bottleneck' [62] [160], that is, the large amount of time and money needed to develop and maintain the formal ontologies. This also includes ontology population and semantic annotation using well-established vocabularies. Existing methods towards semantic annotation and linked data knowledge generation are either a) computationally driven, employing on text mining and information extraction methods or, 2) expert driven. A lack of suitable training data and gold standards make reliability and scalability of computational methods challenging. Expert driven methods involve subject specialists; however, these are often not scalable (time or cost).

2.4 The Need and the Role for Human Computation and Crowdsourcing

To address the challenge of knowledge acquisition bottleneck and others described earlier, researchers have recognized that the realization of the semantic and linked data technologies will require not only computation but also significant human contribution [186]. Emerging research is advocating the use of Human Computation & Crowdsourcing (HC&C) methods to leverage human processing power to solve problems that are still difficult to solve by using solely computers and seek to harness the 'collective intelligence' and the 'wisdom of the crowds' by engaging large number of online contributors to accomplish tasks that cannot yet be automated [97], [200]. HC&C methods are being proposed as well-suited to support the semantic web and linked open data research by providing means to efficiently create research relevant data and potentially solve the bottleneck of knowledge experts and annotators needed for the large-scale deployment of semantic web and linked data technologies [163].

The conceptual framework shown in Figure 2.2 adds and highlights the element and the need of human contribution and intelligence in few important stages such as conceptual knowledge modeling, vocabulary generation, semantic annotation, link generation and verification. Recognizing this need for human intelligence and contribution is fundamental to the motivation behind the core focus of this research i.e. human computation based knowledge engineering. The overall objective is to augment the process of automated knowledge extraction and text mining methods using a hybrid approach for combining collective intelligence of the crowds with that of experts to facilitate activities in formalized knowledge engineering - thus overcoming the so-called knowledge acquisition bottleneck often encountered. As part of this research, it is proposed to build a hybrid methodology of ontology and linked data engineering that combines computational and human computation approaches and attempt to demonstrate the feasibility of this approach when applied to Islamic knowledge. A preliminary prototype version of a basic crowdsourcing workflow, in the context of linked data management only, has been recently published in our research entitled CrowdLink [26], [27].

Follow up ongoing studies by the authors are focusing on case studies that show the amenability of crowdsourcing methods and its variations such as learnersourcing for Islamic knowledge. Some tasks that have been subject to crowdsourcing include: Thematic Disambiguation and Thematic Annotation. Tasks, which have been subject to learnersourcing, include: (1) Thematic Disambiguation, (2) Thematic Annotation, (3) Thematic Classification and (4) Relation Identification. Details may be found in [31] and [25] respectively.

In order to facilitate the process of crowdsourcing and learnersourcing methods, a semantics drive human-machine computation framework has been proposed [27]. This framework provides a workflow mechanism, which allows contributions for the above-mentioned tasks to be obtained from a range of users such as students, teachers and experts. The workflow delegates the tasks according to the difficulty level of the task and skill levels of the users. Simpler tasks are delegated to the regular users or crowd workers and their responses are analyzed. Agreement analytics are applied to filter those responses, which fail to reach a consensus. These responses are then validated by experts using the same workflow framework. Another feature of the framework is tight integration of knowledge extraction and retrieval tasks, tightly embedded in the human computation workflow, which attempt to reduce the human effort. Also, the output of the tasks which are approved, are published as ontologies and linked data. According to the authors, such a framework, which combines human and machine intelligence, is indispensable for achieving the large scale LOD vision for Islamic knowledge outlined in this chapter.

2.5 Conclusions

This chapter presented the emerging vision and the promised potential of Linked Open data for the Islamic knowledge. It has been established that the diverse and distributed body of Islamic knowledge resources present a huge potential for linking, at both macro and micro- level and then publishing this knowledge as linked open data. This would enable significant benefits to the students, teachers, researchers and all kinds of knowledge seekers and educators in general. It would pave way for new kinds of learning applications. However, achieving this vision is a mammoth task, and given the sensitivity of the knowledge at hand, it will not be possible without the intervention of humans. For a scalable approach, this research suggests integrating and automating the role and contribution of humans in knowledge engineering and linked data management processes through the use of human computation and crowdsourcing methods. While the existing research provides evidence of the promises of these methods, adding this dimension to the LOD framework will bring with itself its own set of challenges to the research. However, these very challenges will open doors to various prospective and promising research directions to be undertaken by the Islamic informatics research community at large.

Chapter 3

Human Computation and Crowdsourcing meet the Semantic Web: A Survey

قُلْ إِنَّ صَلَاتِي وَنُسُكِي وَمَحْيَايَ وَمَمَاتِي لِلَّهِ رَبِّ ٱلْعَٰلَمِينَ

Say, "Indeed, my prayer, my rites of sacrifice, my living and my dying are for Allah , Lord of the worlds.

— *[Al-Quran, Al-Anaam,6:162]*

Chapter Overview[1]

In this chapter, we present a comprehensive survey of the intersection of semantic web and the human computation paradigm. We adopt a two fold approach towards understanding this intersection. As the primary focus, we analyze how the semantic web domain has adopted the dimensions of human computation to solve the inherent problems. We present an in-depth analysis of the need for human computation in semantic web tasks such as ontology engineering and linked data management. We provide a 'collective intelligence genome' adapted for the semantic web as means to analyze the threads of composing semantic web applications using human computation methods. As a secondary contribution we also analyze existing research efforts through which the human computation domain has been better served with the use of semantic technologies. We present a comprehensive view of the promises and challenges offered by the successful synergy of semantic web and human computation. In conclusion, we discuss several key outstanding challenges and propose some open research directions.

[1]To Appear:
Basharat, A., I. Budak Arpinar, Rasheed, Khaled, Human Computation and Crowdsourcing meet the Semantic Web, *In the Special Issue on Human Computation and Crowdsourcing (HC&C) in the Context of the Semantic Web, Semantic Web Journal*, 2017.

3.1 Introduction

3.1.1 Research Context

After more than a decade of semantic web research, researchers remain challenged by the large scale adoption of the semantic technologies. Semantic technologies have been deployed in the context of a wide range of information management tasks, for which machine-driven algorithmic techniques aiming at full automation do not reach a level of accuracy and reliability to ensure usable systems. Researchers have started augmenting automatic techniques with human computation capabilities in an effort to solve the inherent problems. Semantic web visionists have put forth a number of visionary ideas for the road ahead for the success of the semantic web. The notion of 'The Global Brain Semantic Web' [34] - a semantic Web interleaving a large number of human and machine computation - has come to be seen as a vision with great potential to overcome some of the issues of the current semantic web . This idea of interleaved human-machine computation has already resulted in successful systems that are able to solve problems in manner and ways unthinkable for either computers or machines to be able to solve alone. The domain of human computation, collective intelligence, social computing and crowdsourcing have all contributed to this successful synergy of humans and machines and contribute to the constantly evolving metaphor of the 'Global Brain' [91]. Much like the concept of the programming needed for the 'Global Brain' [35], the 'Global Brain Semantic Web' will need new strands of programming, workflows and challenges to be accomplished.

The challenge for the semantic web community, is to rethink the original semantic web vision, which was largely built on the vision of computers populating the web of machines [34]. Researchers have recognized the need for human intelligence in the process of semantic content creation [186] which forms the backbone of any semantic application. The entrance barrier for many semantic applications is said to be high, given the dependence on expertise in knowledge engineering, logics and more. In short, semantic web lacks the sufficient user involvement in various aspects. Humans are simply considered indispensable[54] for the semantic web to realize its full potential.

Realizing the potential that human computation, collective intelligence and the fields of the like such as crowdsourcing and social computation have offered, semantic web researchers have attempted to effectively taken up the synergy to solve the bottlenecks of human experts and the needed human contribution in semantic web development processes. Semantic web research can be seen as experiencing a shift from increasingly expert driven to one embracing the larger community and the users involved in the semantic content creation process. Some early efforts that led to the evolution of this approach includes myOntology [187] and inPho [147].

Two major genres of research may be seen emerging in the last few years, in an attempt to bring human computation methods to the semantic web: 1) Mechanized Labour and 2) Games with a Purpose for the Semantic Web. In this chapter, we not only take a detailed look at the challenges leading to the adoption of the human computation methods for the semantic web, we also provide a comprehensive coverage of the approaches in these mentioned genres. On a parallel note, human computation systems can also potentially benefit from the promises offered by the semantic web. The next generation of human computation systems are being envisioned that go beyond the platform they were built on, offering data reusability in ways unintended by their creators [54]. Semantic web may be seen as means of providing better user continuity and platform consistency across human computation systems.

While the potential is clearly evident in going about such a synergy, effectively realizing the synergy of semantic web and human compution will bear its own set of challenges. Both semantic web and human computation seem to have a long way to go before being fully able to reap the benefits promised by the intersection with the other [54].

3.1.2 Contributions

The first contribution of this chapter is to provide a review of the challenges that the semantic web domain has faced especially in terms of the need for human intervention. Secondly, we analyze the intersection of semantic web and the human computation paradigm. We adopt a two fold approach towards understanding this intersection. As the primary focus, we analyze how the semantic web domain has adopted the dimensions of human computation to solve the inherent problems. We present a review of studies and applications spanning the two most common genres in human computation namely Games With A Purpose (GWAP) and Micro-Task Crowdsourcing. We also adopt a classification scheme in the form of 'collective intelligence genome' and apply it to some of the key studies and approaches that combine semantic web, human computation and crowdsourcing. This genome is adapted from the original collective intelligence genome proposed by [127]. We apply it specifically to the context of the semantic web with the aim to provide useful insights in analyzing the various threads that constitute the design of studies for creating semantic web driven by human computation. These threads are considered useful for possible further investigation to be taken up be researchers. At the same time, it is also meant to serve as means of analyzing the strengths and weaknesses of existing researches.

Recent research in crowdsourcing and semantic web has also seen the emergence of some workflow systems designed to meet the need of providing a generic framework for automating human-machine computation workflows. We undertake a comparative analysis of few of the most prominent studies to this end, and highlight the essential dimensions constituting these workflow systems. We give special mention to these systems for they seem closer to the directions that we expect to see in the future that awaits this emerging research domain.

As a secondary contribution, we also analyze existing research efforts through which the human computation domain has been better served with the use of semantic technologies.

We also present a comprehensive view of the promises and challenges offered by the successful synergy of semantic web and human computation.

We hope that this review will serve as basis for exploring newer threads of synergy between the semantic web and human computation research, resulting in the creation of better applications and approaches that advance both domains.

3.2 Background

In this section, we provide a short introduction on some aspects of the theoretical foundations and basics of human computation and crowdsourcing, semantic web, and define concepts that will be used throughout the chapter. We also examine the need of human contribution in the domain of semantic web which provides the grounds for a successful confluence of human computation and the semantic web.

3.2.1 Human Computation and Collective Intelligence: Some Theoretical Foundations

Researchers consider human computation similar to but not synonymous with terms such as collective intelligence, crowdsourcing and social computing. This distinction has been argued in more detail by Quinn and Bederson [156]. While subtle differences may be present, it is evident that these domains are closely knit with one another.

What is Human Computation?

Most widely adopted definition of human computation in state of the art research has been adopted inspired by von Ahn's dissertation titled "Human Computation" [200]. That thesis defines the term as: "...a paradigm for utilizing human processing power to solve problems that computers cannot yet solve". From the body of work that self-identifies as human computation, Quinn and Bederson [156] present a consensus that emerges as to what constitutes human computation: a) The problems fit the general paradigm of computation, and as such might someday be solvable by computers, b) The human participation is directed by the computational system or process.

Classification of Human Computation Systems

With a plethora of contributions from several research domains coming together under the umbrella of human computation, researchers have attempted to distinguish and classify human computation systems highlighting the distinctions and similarities based on several factors such as motivation, quality control, aggregation, process order and task request cardinality [156].

What is Collective Intelligence? Collective intelligence is an encompassing term to broadly refer to groups of individuals doing things collectively that seem intelligent [127, 128]. This idea of loosely organized group of people accomplishing more than what individuals can alone has been garnering a lot of interest within the research community.

What is Social Computing? Technologies such as blogs, wikis, and online communities are examples of social computing [156]. The scope is broad, but always includes humans in a social role where communication is mediated by technology. The purpose is not usually to perform a computation.

What is Crowdsourcing? Crowdsourcing is a term used for a range of activities that take can take different forms as examined by the authors of [64]. The term

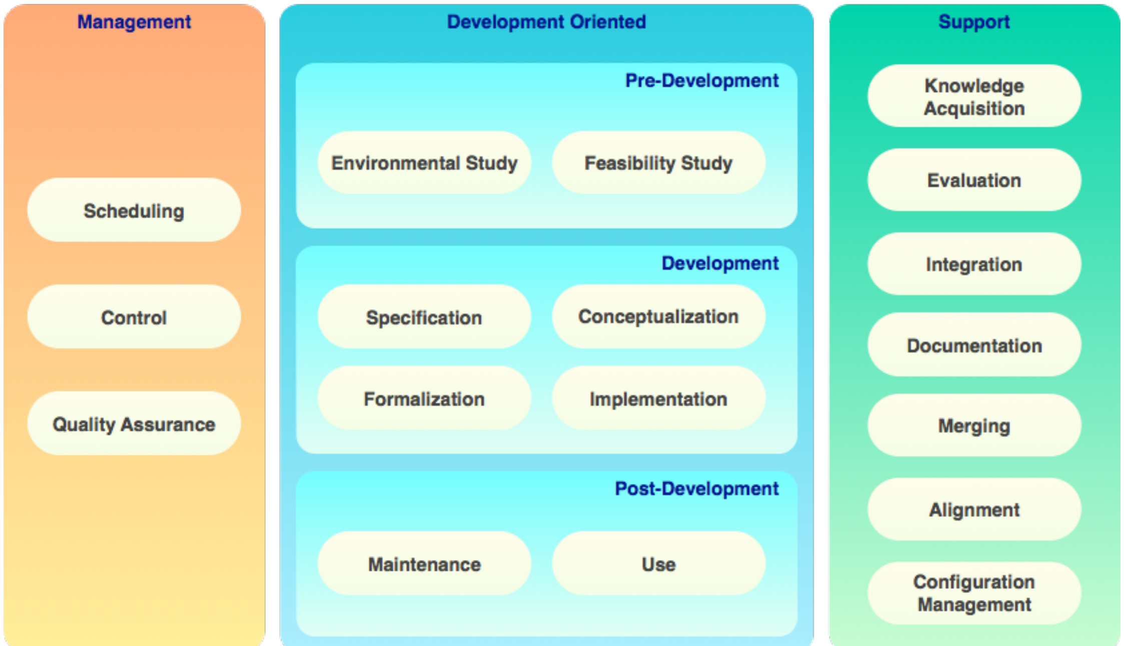


Figure 3.1: Ontology Development Process (simplified adaptation from [153])

was originally coined by Jeff Howe [97, 98]. Essentially, a crowdsourcing system engages a multitude of humans to help solve a wide variety of problems.

It is worth distinguishing between crowdsourcing systems (platforms) vs. crowdsourcing applications. The former refers to those systems such as Amazon Mechanical Turk[2], Turkit [121], CrowdFlower[3], CloudCrowd[4] to name a few that help build crowdsourcing applications in various domains. The latter is a more encompassing term to include any application or system that may incorporate an element of crowdsourcing. Some detailed surveys may be found for further details

[2]www.crowdsource.com
[3]www.crowdflower.com
[4]www.crowdsource.com

[58, 126, 216].

While most obvious forms of crowdsourcing systems engage a crowd of users to explicitly collaborate to build a meaningful and useful artifact, other less obvious forms employ implicit means to gather user contributions such as the 'Games With A Purpose' (GWAP) paradigm. This includes many examples where the users implicitly collaborate. This distinction is provided in the classification by Doan et al. [58].

There are several fundamental issues in crowdsourcing, however, most prominent of thse are: nature of tasks that can be crowdsourced, reliability of crowdsourcing, crowdsourcing workflows. Accoring to Doan et al. [58], crowdsourcing systems face four key challenges: 1) How to recruit contributors, 2) What they can do, 3) How to combine their contributions, and 4) How to manage abuse. In addition, the balance between openness and quality also needs to be ensured. A number of surveys on existing crowdsourcing systems exist [58, 126, 216], which may be referred to for an indepth analysis of the issues involved.

3.2.2 Semantic Web Preliminaries

Tim Berners-Lee envisioned a 'semantic web', capable of providing automated information access based on machine-processable semantics of data and heuristics that utilize this metadata. A new kind of web, driven by explicit representation of the semantics of the data, accompanied with formalized knowledge models, vocabularies in the form of ontologies to enable services operate a greater level of quality [66]. Semantic web was aimed at solving the problems that originally the AI researchers faced in knowledge acquisition, engineering, modeling, representation and reasoning upon knowledge. Given the web enabling a workforce of millions of knowledge 'acquisitioners' working nearly for free providing tons of information and knowledge, semantic web, as the pioneers of semantic web resaerch claimed, was envisioned as the enabling force to build a brain of and for mankind [66], unlike the traditional AI researchers working to rebuild some aspects of the brain.

Ontologies and Ontology Engineering

Ontologies form the backbone of the semantic web. An ontology is an explicit, formal specification of a shared conceptualization of a domain of interest [75, 76]. This famous definition by Gruber implies that the ontology should be machine-readable (formal specification), while at the same time it is accepted by a group or community (shared). Most ontologies are also considered to be restricted to a given domain of interest and therefore model concepts and relations that are relevant to a particular task or application domain [42]. Ontologies are said to formalize the intensional aspects of a domain (also known as the TBox schema), whereas the extensional part is provided by a knowledge base that contains assertions about instances (also known as the ABox schema) of concepts and relations as defined by the ontology.

Understanding the Ontology Development and Knowledge Engineering Processes:

Figure 3.1 gives an overview of the activities involved in the *Ontology Devel-*

opment process adapted from [153] . This process clearly distinguishes between three key groups of activities namely *Management*, *Development* and *Support*.

Ontology Management: This refers to primarily the activities of scheduling, controlling and quality assurance. Scheduling involves activities that pertain to the coordination and management of an ontology development project such as resource and time management. Controlling deals with ensuring that the scheduled tasks are accomplished as planned. Quality assurance process is involved in evaluating the quality of the outcomes of each activity, in particular of the developed ontology.

Ontology Development: This is further divided into three phases: pre-development, development and post-development. In the pre-development phase, an environmental study investigates the intended usage and the competencies of the envisioned ontology. The feasibility study ensures that the ontology can actually be built within the time and resources assigned to the project. The development stage of ontology development includes specification, which defines the usage, users and the scope of the ontology. This is usually followed by conceptualization in which domain knowledge is structured into meaningful models, followed by formalization and implementation. During formalization, the conceptual model is formalized to prepare the implementation in a given ontology language. In the post-development phase, the ontology is subjected to use, updates, and maintenance as required.

Ontology Support: This covers a broad range of activities which are considered to be most crucial and cover many areas of semantic content creation and maintenance. No order is determined in which these activities must be undertaken. Typical support activities include knowledge acquisition (specification of the knowledge needed for ontology), learning (process of automatically creating an ontology or a part of it), evaluation (analyzing the quality of the developed ontology), integration (for reusing other ontologies to build a new ontology), documentation (for providing detailed description of the ontology and main activities and tasks of the development process), merging (produces a new ontology by combining existing ontologies), configuration management (tracks versions of the ontology) and alignment (establishes relations among related or complementary ontologies).

Researchers have primarily focused on the development and support activities because they are specific to the scope of the semantic content creation.

Linked Open Data

Linked Open Data (LOD) has recently been emerged as 'a set of best practices for publishing and connecting structured data on the Web' [38]. Linked data initiative claims to provide two primary advantages by exposing data in the form of linked data. These are data discovery and data integration at web-scale, in a uniform and homogenous manner. Linked open data is based on two simple ideas [89] : First, to employ RDF to publish structured data sources. Second, to explicitly interlink different data sources. Tremendous amount of data is published on the Web according to the linked data principles [38, 105]. LOD has generated remarkable threads of arguments among research communities. The number of participants in the LOD cloud has been significantly growing and publishing, consuming and reuse of structured data on the web has been altered dramatically. Apparently, managing data in such highly distributed and heterogeneous environment is a critical issue and hence has opened up many research opportunities. We examine the issues that require special input from humans in the next sub-section, thus making LOD management a key area subject to human computation.

3.2.3 Human contribution in the process of semantic content creation

We focus on two strands of semantic content creation: 1) Ontology Development and 2) Linked Data Management. We comprehensively analyze and summarize the need for human intelligence in contribution in both these contexts.

Need for Human Contribution in the Ontology Engineering Process

In order to highlight the role of human intelligence in semantic web research, Siorpaes and Simperl [186] surveyed methodologies, methods and tools covering various activities in the ontology engineering lifecycle in order to learn about the types of processes knowledge and ontology engineering projects conform to, and the extent and reasons they might rely on human intervention. We summarize their findings here with an overview shown in Figure 3.2 as derived from their analysis. They analyze and classify human contribution falling in three key stages of Ontology Engineering stages described in Section 3.2.2 earlier. These are: Ontology Development, Semantic Annotation and Ontology Support. In the figure, the broad task categories are classified according to the level of automation and the extent of human contribution usually required by these activities.

Ontology Development: A further detailed breakdown of ontology development tasks is given in Figure 3.3. It is obvious, that pretty much all tasks require human intervention, while only few are supported with automation support. Conceptual modeling, which is one of the most essential tasks in ontology development, is essentially a human-driven task. Only very few tasks can be automated and the final modeling decision is always taken by human actors. Automation support is partly possible for collecting relevant terms or for detecting properties of concepts based on semi-structured knowledge corpora such as folksonomies.

Semantic Annotation Automation has more role in the annotation of several types of media such as text annotation that leverages on natural language processing techniques and has been investigated upon in a plethora of research. However, human supervision is needed to ensure reliability. The annotation of multimedia content is more challenging, whereby only low level semantic content may be automatically derived. Much of the high level semantics is derived using human effort. The annotation of web services is very much manual task.

Ontology Learning: Although, it supported with tools that may be automatically executed, require human effort in terms of input and feedback. A variety of tools are available ranging from fully automatic, to semi-automatic, to those that heavily rely on user intervention through out the process.

Ontology Alignment: This is another activity that is said to be specific in nature and far from a stage where any successful generic methodologies have been seen. While there are numerous tools to match, merge, integrate, map and align heterogeneous ontologies, very few tools can be run fully automatically. Most tools require a human in the loop, to give feedback or input suggestions to the system. Some even depend on user defined rules to carry out the mappings.

Ontology Evaluation: This is a broad topic, which is nevertheless primarily driven by human effort. Automatic support is available to cross-check and validate structural features of taxonomies. Some tools are designed to facilitate the humans to carry out the evaluation. The notion of achieving an automated process of ontology evaluation generic enough to be applied across domains is hardly feasible, given the depth of knowledge needed that is difficult to be captured in tools that run on their own.

Need for Human Contribution in Linked Data Management

Linked data, fueled primarily by the semantic web vision, is considered to be one of most important technological developments in the data management perspective in the last decade [3]. However the seamless consumption and integration of linked open data is challenged by the several quality issues and problems that the linked data paradigm is facing. As researchers remark, many of these quality issues are not possible to be fixed automatically rather, require manual human effort.

Emerging research is establishing and highlighting the need to combine human and computational intelligence, instead of relying on fully automated approaches. Simperl et al., [179] argue that several tasks in the context of linked open data are inherently human driven, given their highly contextual and often knowledge intensive nature, which imposes considerable challenges to the the algorithmic approaches. They provide a broad classification of tasks derived from the study

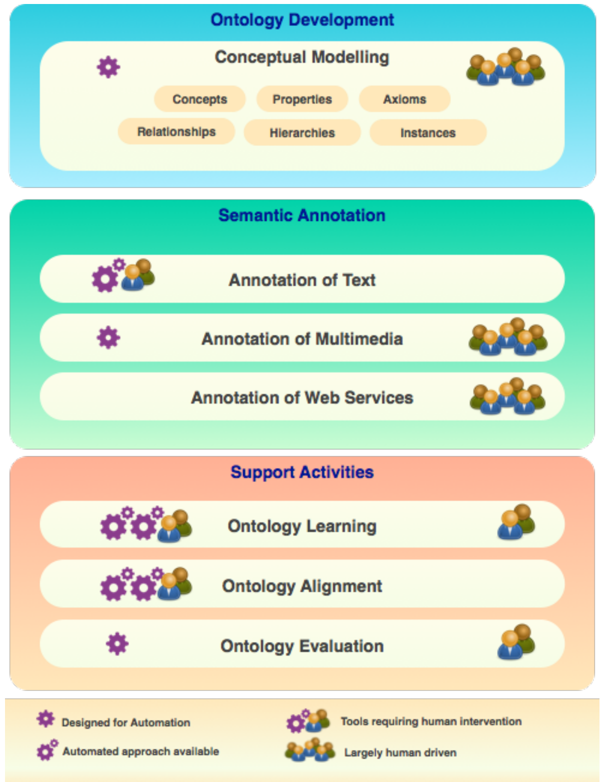


Figure 3.2: Classification of Tasks in the Ontology Engineering Process (as summarized in [186])

of the architecture for applications that consume the linked data as suggested in [89]. [3], [4] and [26, 27] also provide insights into the tasks management for the linked data from the human intelligence perspective.

We summarize the broad group of tasks that could be subject to semi-automation and human computation based on what has been identified by the existing studies, and that are relevant to the linked data architecture.

Figure 3.3: Classification of Tasks in the Process of Ontology Development (as summarized in [186])

Linked Data Annotation and Production:

- *Identity Resolution*: This involves creating *owl: sameAs* links between different entities, either by comparison of metadata or by the investigation of links on the human web.

- *Entity Linking*: This may involve linking two entities using a new relation between them. This is different from identity resolution since the new link may not always be an *owl: sameAs* relation.

- *Entity Classification*: Unlike the emphasis of the traditional semantic web ontologies, the LOD tends to emphasize the relationships and links between the entities, rather than classification of entities. Since the vocabularies used are generic, it is difficult to infer a classification using the established reasoning methods and techniques. Therefore this type of classification becomes dependent on human contribution.

- *Ordering or Ranking Facts*: Often the linked open data facts need to be ranked depending on the needs for querying and browsing the data. This is again a task that is difficult to be automated and would require human intervention.

- *Generating new facts*: Linked data suffers greatly from the lack of completion of information. For instance, a common need is for labelling in terms of multilinguality. Providing non-english labels (Translation) is something human contribution can not only benefit from, rather it is indispensable in certain situations.

- *Creating Links with the LOD*: Applications consuming the LOD will often attempt to create links with the entities in the LOD datasets. This too cannot be automated entirely and often requires human input.

Quality Assessment of Linked Data: This includes the following:

- *Validation of Linked Data*: Many aspects of linked data need manual validation in terms of completion, correctness and provenance.
- *Meta-data Completion*: At times the extraction process used to create linked data results in the incomplete extraction of entities from the source datasets, resulting in incomplete meta data information, which needs to be manually inspected, identified and corrected.
- *Meta-Data Correction*: It is also possible that the meta-data information is extracted incorrectly. This is something that is difficult to be detected automatically and requires human intervention for identification and correction.
- *Link Evaluation*: Links on the LOD often need to be manually inspected, for correctness and verification, even though automated tools may help in identifying some problem areas. At times the link, though valid, doesnot show any valid content against the linked entity. Such problems are hard to be identified and corrected in an automated manner.

Query Processing: Since any form of consumption of linked open data requires some form of query processing, primarily carried out using the SPARQL language, it is natural that this task will involve human intelligence at various levels.

We will revisit these tasks with details of how they have been subjected to human computation methods, in particular crowdsourcing in Section 3.3.2.

3.2.4 Confluence of Semantic Web and Human Computation

After more than a decade of semantic web research, researchers remain challenged by the large scale adoption of the semantic technologies. The question of why and where the element of human computation is needed in semantic web has been answered in the previous sub-section.

Simperl et al. [182] argue that human computation methods such as paid micro-tasks and games with a purpose could be used to advance the state of the art in knowledge engineering research in general and ontology engineering in particular. They examine the literature to note how human computation has received great attention from the semantic web research community, resulting in a number of systems that tackle essentially human driven tasks that require expertise such as ontology classification, annotation, labelling, ontology alignment and annotation of different types of media to name a few.

Looking into the reasons of barriers imposed by the adoption of semantic technologies, one finds that there is considerable lack of useful semantic content as this content cannot be created automatically but requires to a significant degree, human contribution. Ironically, there is not enough interest in creating semantic content from the humans. Abraham Bernstein, in his recent vision towards the 'Global Brain Semantic Web', highlights important differences between human computers and traditional computers [34]. He enlists three major differences: Motivational Diversity, Cognitive Diversity and Error Diversity. These diversities give rise to a range of issues when including people in semantic web.

Research clearly indicates that combining human computation and semantic web is of mutual benefit to both domains i.e. its benefits are not uni-directional. Both domains are complementary to one another. Although, the benefits of one may have superseded the other in terms of the current landscape of research. More research is evident in terms of how semantic web domain has benefitted from human computation. More real examples of systems whereby the semantic web has driven the human computation system towards a better future are yet to be seen. The future of human computation and semantic web holds great promise according to Difranzo and Hendler [54] who claim, while remarking on the mutual promise that the two fields offer, that human computation can be used to help curate the semantic web data in the linked open data cloud while the semantic web can be used to provide better user continuity and platform consistency across human computation systems. However, many challenges and questions still remain. Leveraging the best of human computation and semantic web technologies is greatly needed to bring forth this next generation.

3.3 Application of Human Computation Techniques to Semantic Web Research

In this section, we first provide an overview of those human computation genres that have been applied to the semantic web research. We then present a classification of the broad tasks and applications in the semantic web domain according to these genres. We also review tool support for closely integrating human computation into semantic web processes.

3.3.1 Human Computation Genres for the Semantic Web

Applications of human computation techniques to the domain of semantic web falls primarily under two genres: 1) Games with a Purpose (GWAP) and 2) Mechanized Labor or Micro-task Crowdsourcing. We present a brief overview of both approaches and then present classification of existing approaches according to the two genres.

Semantic GWAPs: Semantic Games With A Purpose

The philosophy that GWAPs capitalize upon is simple yet elegantly effective: tasks that are difficult for the computers but can be tackled easily by the humans are hidden behind entertaining games. The games are designed to engage regular users and not just experts. Users are usually unaware when they are playing such games how they are indirectly helping build knowledge bases and create annotations useful for further computation and to be used in algorithms. This is a significant step in combining human and machine intelligence. Siorpaes and Hepp [183] adopted the Lui von Ahn's "games with a purpose" [199, 201, 202] paradigm for creating the next generation of the semantic web. They proposed the idea of tapping on the wisdom of the crowds by providing motivation in terms of fun and intellectual challenge. According to von Ahn, the key principle in GWAPS is the idea that people are not interested in solving a computational problem, rather they want to be engaged in something entertaining. The primary objective of semantic GWAPS is massive semantic content generation which entails that there will be massive user participation as well. Pe-Than et al. [152] review human computation games in general and provide a typology consisting of 12 dimensions and strategies employed by the games. Simko and M. Bielikova [173] present a classification of semantics acquisition games' purposes, ranging from multimedia metadata acquisition, through text annotation, building common knowledge bases to ontology creation. Simko and his collegues have a series of work which may be refered to with respect to semantics discovery and acquisition using human computation games [171, 204].

Recent work by Simperl et al., [176] have proposed an ontology learning approach using games with a purpose.

Micro-Task Crowdsourcing

Several recent research prototypes have attempted to use micro-task crowdsourcing for solving semantic web tasks. Some of the most common platforms used for crowdsourcing include the Amazon Mechanical Turk[5](referred alternatively as AMT or MTurk) and CrowdFlower[6], amongst others. MTurk (and others) provides a platform where users (*requesters*) can post a given problem(*task*) that other users (*turkers*) can solve. To do so, the requester designs a problem, to a suitable degree of granularity, resulting in a number of so-called Human Intelligence Tasks (HITs), which can be tackled independently by multiple turkers in return for a financial reward.

[5] www.mturk.com
[6] www.crowdflower.com

Figure 3.4: The Typical Task Lifecycle in a Crowdsourcing Platform

The complete execution of a task (HIT) lifecycle using any particular crowdsourcing system involves several general stages, as shown in Figure 3.4: (i) A task is defined, with input parameters and the needed data, (ii) The task is published over the crowdsourcing platform, (iii) The task is searched for or discovered by the crowdworkers (via search or direct notification), (iv) The crowdworkers perform and submit the tasks, (v) The submitted tasks then need to go through a review process for obtaining the submissions, (vi) The submissions are either accepted or rejected and (vii) The accepted submissions are further processed for retrieving and synthesizing relevant results.

There may be a number of additional configuration parameters when tasks are published or submitted such as the number of assignments, the time to completion for each HIT or restrictions on the profiles of the workers. Once the tasks are completed, and submitted the results are collected and aggregated and quality assurance measures applied depending on the the nature of the task design. The evaluation may be carried out in a number of ways. Often, a common practice to allow assignments of the same task to multiple workers. Therefore the results may be aggregated using majority voting or other sophisticated techniques such as a probability distribution or by taking into account some estimate of the expertise and skills of the works.

An important conceptual distinction has to be made between the terms tasks, micro-tasks and complex tasks. This has been done by Luz et. al [126]. Micro-tasks are atomoic operations, whereas complex tasks are sets of microtasks (e.g some workflow) with a specific purpose. In this section, we adopt a generic approach to crowdsourcing approaches inclusive of both types of tasks. We dedicate a separate discussion to those systems that involve some form of workflow based approach as the basis of their crowdsourcing model in Section 3.5.

There are several factors known from existing research which constrain the applicability of a micro-task appraoch e.g. according to [179] there are three key factors:

Decomposability: The tasks need to be decomposable into appropriate levels of granularity which can be executed independently. This also needs to be done according to the platform upon which the task is being executed e.g. AMT. If the granularity of the task is higher than the level at which it can be atomically published, then additional workflows need to be generated and managed such as the approach defined in [115].

Verifiability: The performance of the workers need to be measurable. This entails the need for methods and techniques for the quality assessment of the collected results. In addition, mapping to the input from different task assignments are also needed. Open-ended tasks such as requiring the definition of a term or translation requires means to deal with iteration such as one discussed in [121, 123].

Expertise: The domain of the tasks being experimented ought to be available to the workers. However, some knowledge intensive tasks may require additional expertise which the general purpose platforms such as AMT are not expected to provide.

Hybrid Genres: Combining GWAPs and Micro-Task Crowdsourcing

There is evidence of research that combines the two human computation genres in an attempt to cache upon the relative benefits of both approaches. Thaler et. al [194] compared these two prominent techniques to evaluate which approach is better with respect to costs and benefits. They used OntoPronto GWAP and replicated the study by employing the Amazon Mechanical Turk using a similar approach. Their experimentation showed the feasibility of both approaches in accomplishing conceptual modeling tasks, however, they concluded the microtask approach to be superior in terms of both development and problem solving tasks. There are tradeoffs to both approaches. There is also evidence of using a hybrid-genre workflow [160] that seeks to combine the two genres in a single workflow to achieve some promising results. However, the hybrid approach may not be feasible in all kinds of scenarios and needs more dwelling into further.

Another example of a hybrid genre is the combination of a 'contest' and a paid micro-task approach used by Acosta et al. [3]. The contest targets the experts and linked data enthusiasts, whereas the paid microtasks target faceless crowds on the Amazon mechanical turk. The study is an interesting contribution to show how such a hybrid methodology, whereby both approaches used not only complement one another, they may be optimally combined together to enhance linked data curation processes.

Task Category	**Task Description**	**Genre**	**Source**
	Ontology Engineering: Ontology Creation/Ontology Alignment		
Concept and Instance Identification	Decide if given entity forms a class or an instance	GWAP	OntoPronto [183] OntoGame [184]
	Identify class from given attributes Given a set of attribute descriptions, identify the class	GWAP	GuessWhat?! [130]
	Validate class labels Evaluate if the class name fits the description	GWAP	GuessWhat?! [130]
Specification of Term Relatedness	Check whether two terms (usually ontology concepts) are equal Select from term pairs	MTurk	InPhO [62]
	Select from a set of terms	GWAP	SpotTheLink [191, 192]
	Voting on terms	GWAP	LittleGameSearch [174]
	Provide related terms	GWAP	FreeAssociation [196]

Verification of Relation Correctness **Verification of a valid or invalid relation**	Equivalence	CrowdFlower	CrowdMap [164]
		MTurk	Conference v2.0 [46]
	Subsumption	MTurk	Noy et al. [148]
		CrowdFlower	CrowdMap [164]
Specification of Relation Type **Select an appropriate relation from the set of given relation types**	Equivalence	CrowdFlower	CrowdMap[164]
			SpotTheLink [191, 192] OntoPronto[183]
	Subsumption	MTurk	InPhO [62]
		CrowdFlower	CrowdMap [164]
		GWAP	SpotTheLink [191, 192] Categorilla [196] OntoPronto [183]
	Instantiation (Specify Instance Types)	GWAP	Categorilla [196]
	Disjointness	GWAP	OntoPronto [183]
	InstanceOf	GWAP	OntoPronto [183]
	Spatial, Purpose, Opposite of, IsRelated to	GWAP	Verbosity [203]
	Generic and Domain Specific Relations	GWAP	ClimateQuiz [165]
Ontology Engineering: Relevant for Ontology Learning/Automatic Extraction of Ontologies			
Verification of Domain Relevance	Verify if the given term is relevant to the domain	GWAP	OntoGalaxy [114]
Annotation of Multimedia	Select from given choices and annotate a video	GWAP	OntoTube [183]
		CrowdFlower, MTurk	CrowdTruth [100]
Extraction of Text Annotations	Medical Text Annotation	GWAP	Dr. Detective [61]
	Generic text annotation	CrowdFlower, MTurk	CrowdTruth [100]
Annotating Web content	eBay offereings	GWAP	OntoBay [183]
Domain Specific Vocabulary and Relation Building	Collect goals and attributes	GWAP	Common Consensus [119]
	Collect terms and relationships	GWAP	TermBlaster [172]

Table 3.1: Broad Classification of Human Computation Tasks in the Semantic Web (Ontology Engineering) and the Respective Genres

3.3.2 Classification of Human Computation Tasks and Genres in the Semantic Web

A broad classification of tasks subject to human computation in the semantic web domain and respective genres is given Figures 3.5 and 3.6 and detailed classification of related research and applications is presented in Tables 3.1 and 3.2. The broad classification is adapted from the reviews presented in [5, 54, 182, 186] and a study of the systems and applications enlisted. The classification is broadly classified in the categories of ontology engineering and linked data management tasks. The classification of the studies show that the role of GWAPs in the ontology engineering tasks is significantly greater than those of mechanized labor and there is considerable room for more research.

Ontology Engineering

Research in ontology engineering not only recognizes and establishes the need for human intelligence and contribution [186], there have been considerable efforts to develop ontologies in a collaborative and community driven manner. A detailed review may be found at [178]. CrowdMap [164] attempt to solve ontology alignment tasks using CrowdFlower using a workflow framework designed to crowdsource ontology alignment tasks. Mortensen et al., [140, 141, 142, 143, 144] and Noy and collegues [148] present a series of works with respect to ontology validation and quality assurance in the domain of biomedical ontologies. The work serves to establish strong grounds for the feasibility of crowdsourcing in both generic and specialized domains. A broad overview of these tasks is presented in Figure 3.5.

A detailed classification of ontology engineering tasks using human computation in presented in Table 3.1.

- *Concept and Instance Identification/Collection*: One of the most fundamental tasks in Ontology engineering is the process of identifying and collecting concepts and instances. Ontogame [184] and OntoPronto [183] adopt a game based approach to decide if given entity forms a class or an instant. Players are presented with a Wikipedia page of an entity and they have to judge if it denotes a class or an instance and then relate it to the most specific concept of the PROTON ontology. GuessWhat?! [130] also includes the similar feature, whereby, using a GWAP, the player is supposed to identify class from the given attributes and description. A similar task also incorporated by GuessWhat?! [130] is that of validating class labels, in which the user is required to evaluate if the class name fits the description.

- *Specification of Term Relatedness*: In this particular type of task, the crowd workers check whether two terms, which are usually concepts in an ontology, are equal. InPhO [62] is an example for such a study which employs MTurk for the purpose of this task, whereby workers select from a choice of term pairs given to them to select the pair that represent the equality in the best manner. Workers may be required to select from a set of given terms (SpotTheLink [191, 192]), or vote on terms (LittleGameSearch [174]). In some cases, workers or players provdie related terms (FreeAssociation [196]).

- *Verification of Relation Correctness*: In this task, crowd workers or players are usually presented with a pair of terms and a relation between these terms and they are required to determine if the relation is a valid or an applicable one. Most common of these relation types include equivalence (CrowdMap [164], Conference v2.0 [46]) and subsumption (Noy et al. [148], CrowdMap[164]).

- *Specification of Relation Type*: This task focus on the selection of an appropriate relation from the set of given relation types. Most common relation types that have been crowdsourced so far are equivalence (CrowdMap [164], SpotTheLink [191, 192], OntoPronto [183]), subsumption (CrowdMap[164], InPhO [62]), instantiation (Categorilla [196]), disjointness (OntoPronto[183]) and instanceOf (OntoPronto[183]). There are other examples such as spatial, purpose, oppositeOf, isRelatedTo provided by Verbosity [203]. There are also efforts to crowdsource domain specific relations such as ClimateQuiz [165].

- *Verification of Domain Relevance* : Some tasks require the crowdworkers to verify if the given term is relevant to a specific domain such as OntoGalaxy [114].

- *Annotation of Text and Multimedia*: Several studies include the annotation of text (Crowd-Truth [100]) and multimedia (OntoTube[183], CrowdTruth [100]).

- *Annotation of Web Content*: Another aspect of semantic annotation is the annotation of the web content. OntoBay [183] for instance gets annotations

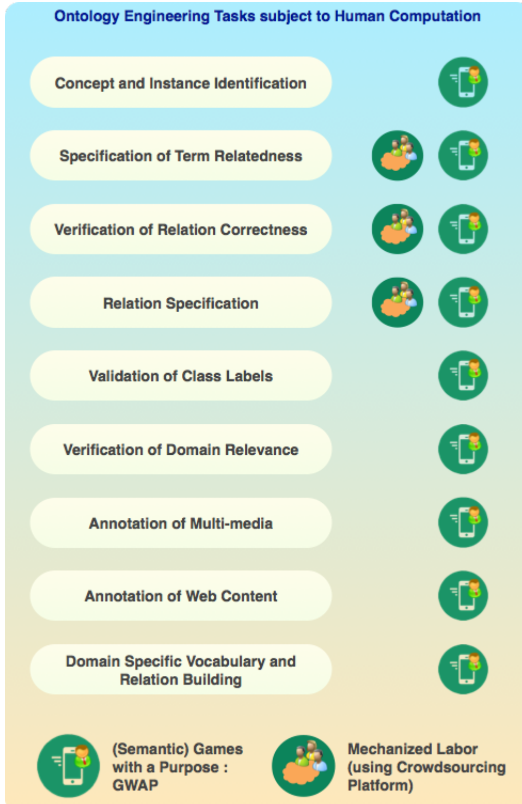


Figure 3.5: Classification of Human Computation Tasks and Genres in Semantic Web (Ontology Engineering)

for eBayOfferings.

- *Domain Specific Vocabulary and Relation Building*: There also exists evidence of studies that work on domain specific relation and vocabulary building. Common Consensus [119], for instance, collects goals and attributes while TermBlaster [172] collects terms and relationships between them.

It is obvious from Table 3.1 and Figure 3.5 that a wide range of tasks have been subjected to human contribution using both mechanized labor and GWAPs. While the classification is not meant to be exhaustive, however, it shows representative studies in broadly most categories of ontology engineering tasks reported in literature. It is also evident, that interestingly, much of ontology engineering tasks have so far been tackled using the GWAP approach, implying that there is much potential and room for the mechanized labor to be taken up further towards a holistic ontology engineering approach.

Task Category	Task Description	Genre	Source
	Linked Data Management		
Linked Data Annotation and Production			
Identity Resolution	Mapping entity links	Crowd (MTurk)	Simperl et al. [179]
	Entity/Link disambiguation	Crowd (MTurk)	CrowdLink [26, 27]
Entity Linking	Matching entities to most similar ones by picking links (Matching URIs)	MTurk	ZenCrowd [51, 52]
Entity Classification	Classification	Crowd (MTurk)	Simperl et al. [179]
Ranking Facts in LOD	Ordering	Crowd (MTurk)	Simperl et al. [179]
	Play Jeopardy like quiz game to rank facts	GWAP	RISQ! [210]
Generating New Facts	Translation	Crowd (MTurk)	Simperl et al. [179]
	Collaborative image annotation	GWAP	SeaFish [193]
	Create new facts	Crowd (MTurk)	CrowdLink [26, 27]
	Linguistic linked open data Production Play word game and add word senses and translations	mGWAP	WordBucket [159]
	Generating ground truth by anwering questions on a quiz	GWAP	WhoKnows?Movies! [195]
Create Links with LOD Concepts	Link photographs with LOD concepts	mGWAP	UrbanMatch [44, 45]
Quality Assessment of Linked Data			
Validation of Linked Data	Validating links	Crowd (MTurk)	CrowdLink [26, 27]
	Review concepts in DBPedia while reviewing restaurant reviews	App	Taste It! Try It! [185]
	Link verification	GWAP	VeriLinks [117]
Metadata-Correction	Metadata-Correction	Crowd (MTurk)	Simperl et al. [179]

	Incorrect data type	Contest (TCM Tool) & Crowd (MTurk)	Acosta et al. [3]
Metadata-Completion	Incomplete object value	Contest (TCM Tool) & Crowd (MTurk)	Acosta et al. [3]
Link Evaluation	Incorrect links	Contest (TCM Tool) & Crowd (MTurk)	Acosta et al. [3]
	Link verification	MTurk	CrowdLink [26, 27]
	Evaluating LOD heuristics using a quiz	GWAP	WhoKnows? [205]
Query Processing	Query processing	MTurk	CrowdSPARQL [4, 180]

Table 3.2: Broad Classification of Human Computation Tasks in Semantic Web (Linked Data Management) and Respective Genres

Linked Data Management

The Linked Data initiative has come to be seen as a significant step forward towards the large scale adoption of the semantic web. However, the linked data management, which includes linked data creation, updation, querying, retrieval, validation and quality assessment itself presents its own challenges. Linked data researchers have also capitalized upon the benefits of crowdsourcing, GWAPS and human computation to solve several challenges [185]. Simperl and Wolger et al., [181] provide a survey of various data interlinking tools in order to highlight aspects of the interlinking process that crucially rely on human contributions and explain how these aspects could be subject to a semantically enabled human computation architecture that can be set-up by extending interlinking platforms such as Silk with direct interfaces to popular microtask platforms such as Amazon's Mechanical Turk.

Linked Data Annotation and Production Tasks:

- *Identity Resolution*: This involves asking the crowd to help create *owl:*

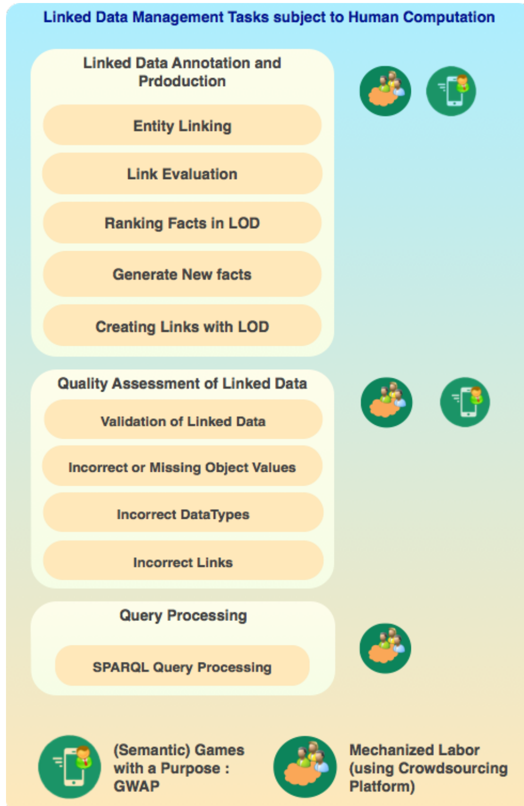


Figure 3.6: Classification of Human Computation Tasks and Genres in Semantic Web (Linked Data Management)

sameAs links between different entities, either by comparison of metadata or by the investigation of links on the human web. Simperl et al. [179] use the MTurk to map entity links. A similar approach for entity disambiguation has been adopted by CrowdLink [26, 27] whereby the crowd disambiguates the researchers DBLP names and links using the information provided.

- *Entity Linking*: This task requires the crowd to link two entities using a new relation between them. This is different from identity resolution since the new link may not always be an *owl: sameAs* relation. For instance, ZenCrowd [51, 52] utilizes MTurk to involve the crowd in matching entities to most similar ones by picking links (Matching URIs).

- *Entity Classification*: Simperl et al. [179] propose the approach of entity classification which involves classifying entities to pre-determined vocabularies or ontologies.

- *Ordering or Ranking Facts*: Often the linked open data facts need to be ranked depending on the needs for querying and browsing the data as proposed by Simperl et al. [179]. RISQ! [210] uses a GWAP approach to rank facts by asking players to play a jeopardy like quiz game.

- *Generating new facts*: Linked data suffers greatly from the lack of completion of information. E.g. a common need is for labelling in terms of multilinguality. Providing non-english labels (translation) is something human contribution can benefit from as proposed by Simperl et al. [179]. SeaFish [193] uses a GWAP approach for creating annotations for images in a collaborative manner. CrowdLink [26, 27] provides means to create or acquire new knowledge and facts for LOD sources using MTurk. WordBucket [159] uses a mobile based GWAP approach and attempts to create linked data by asking the players to play word games and thus adding word senses and translations. WhoKnows?Movies! [195] also uses a GWAP approach to generate ground truth by requiring the players to answer questions on a Quiz.

- *Creating Links with the LOD*: Applications consuming the LOD will often attempt to create links with the entities in the LOD datasets. UrbanMatch [44, 45], for instance, uses a mobile GWAP approach to link photographs with LOD concepts.

Quality Assessment of Linked Data: When it comes to quality assessment whether semantic web in general or linked open data in particular, the fitness-for-use principle [109] demands the involvement of humans to ensure such quality.

- *Validation of Linked Data*: Many aspects of linked data need manual validation in terms of completion, correctness and provenance. In this respect, CrowdLink [26, 27] provides a crowdsourcing based workflow for verifying entity links from LOD sources. Taste It! Try It! [185] is an application that aims to get users to review concepts in DBPedia while reviewing restaurant reviews. VeriLinks [117] is another example that uses GWAP approach for link verification. A relevant issue when it comes to linked data curation using human computation is the issue of data provenance and version control on linked data resources. A collaborative approach to curate linked data has been proposed by Knuth et al., [111]. At the core of their approach lies the design and usage of the PatchR ontology [110] that allows to describe patch requests including provenance information. They extended WhoKnows? [205] to embed publishing prevenance information within a GWAP. Markovic et al, examine the role of provenance in social computation [134], and demonstrate the importance and management of provenance of crowdsourced disruption reports [134]. The proposed vocabulary SC-PROV, a provenance vocabulary for social computation [135] is an important step in managing the provenance information and thus contributing to enhancing the transparency in human computation systems. This is crucial for enabling decision making about the reliability of participants and quality of the generated solutions.

- *Meta-Data Correction*: For identification and correction of incorrecte meta-data, Simperl et al. [179] and CrowdLink [26, 27] provide MTurk based tasks which can easily post tasks to the crowd.

- *Meta-data Completion*: Acosta et al. [3] provide means to verify and complete incomplete meta data information which needs to be manually inspected, identified and corrected. They provide both a MTurk approach and a custom tool called TripleCheckMate.

- *Link Evaluation*: Acosta et al. [3] use their tool TripleCheckMate to identify incorrect link and also use MTurk to do the same. CrowdLink [26, 27] also provide means for link verification using their crowdsourcing framework for linked data management. WhoKnows? [205] uses a GWAP approach for evaluating LOD heuristics using a quiz.

Query Processing Tasks:

CrowdSPARQL [4, 180] is an approach proposed for carrying out a number of linked data management tasks using SPARQL query processing, and delegating part of the triples in the query to the crowd using MTurk. This is a novel approach indeed. There are similar successful approaches in the database community such as CrowdDB [69].

3.3.3 Automated Tool Support for Human Computation based Semantic Web Tasks

Most applications described in Section 3.3.2 and 3.3.2 are primarily independent, external systems or applications. There are relatively fewer efforts that aim to combine and closely integrate crowdsourcing into ontology engineering practices. We briefly review such efforts in both ontology engineering and LOD management perspective.

Automated Tool support for crowdsourcing Ontology Engineering Tasks: To reduce the complexity of ontology construction, the process is often bootstrapped by re-using existing or automatically derived ontologies. Ontology learning methods are often used to automatically extract ontologies from structured and unstructured sources. Crowdsourcing is therefore seen as a useful method for seeking human contribution in various ontology engineering tasks. However, despite the usefulness, the process of acquiring crowdsourced inputs brings its own baggage. Hanika et al, [81] on remarking the need for automated tool based support, highlight the high upfront investments(understanding techniques, task design) in setting up crowdsourcing for ontology engineers.

Researchers in the domain of Natural Language Processing (NLP), have adopted crowdsourcing techniques [161]. A recent effort has successfully engineereed a crowdsourcing plugin called GATE Crowdsourcing plugin [39], as a new component in the populate GATE NLP system that allows seamless integration of crowdsourcing tasks into larger NLP workflows, from within the GATE's user's interface. Noy et al, [148] have also envisioned a similar tool to support ontology engineering tasks. Inspired from these ideas and efforts, the foremost effort to this end, is the *uComp Protégé plugin* that aims to closely integrate typical crowdsourcing tasks into the ontology engineering work from within the Protégé ontology editing environment [5, 80, 81] .

The plugin allows for a number of ontology engineering tasks to be crowdsourced. The main stages involved when using the plugin are shown in Figure 3.7. The efficiency of the plugin is compared to manual efforts in terms of time and cost reductions, while ensuring good data quality. The findings indicate that in scenarios where automatically extracted ontologies are verified and pruned, the use of the plugin significantly reduces the time spent by the ontology engineer and leads to important cost reductions without a loss of quality with respect to manual process.

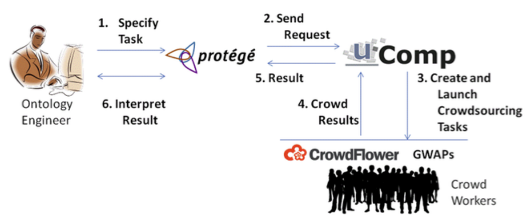


Figure 3.7: Main Stages in using the uComp Plugin from [5]

Automated Tool support for crowdsourcing Linked Data Tasks: Automated tool support for Linked data management has been proposed using the crowd by

TripleCheckMate: a tool for crowdsourcing the quality assessment of linked data [113]. The tool has been successfully used to crowdsource issues of accuracy, relevancy, representational consistency and interlinking quality dimensions (a subset of those described in [219]) applied to DBPedia.

3.4 A Collective Intelligence Genome for the Semantic Web

After having analyzed the classification of broad semantic web engineering tasks and the respective human computation genres, we present the collective intelligence genome for the semantic web, which aims to highlight the essential threads that constitute the successful design considerations for conceiving an application that successfully integrates human computation.

3.4.1 Overview

Malone et al. in their seminal work present a user's guide to building blocks of Collective Intelligence (CI) [127, 128] . They try to answer the fundamental question of how to get the crowds to do what needs to be done. They claim that combining or recombining the right *genes* or building blocks would result in the formulation of the right kind of system suited to the needs of collective intelligence. We adopt this genome and apply it to the existing research limiting our focus to primarily those studies that attempt to apply collective intelligence, human computation or crowdsourcing of some sort in the context of semantic web research. We analyze how well these studies comply with this genome and identify any possible gaps, limitations and opportunities in this light.

3.4.2 Key Driving Questions

The key driving questions that drive the CI genome design are: What is being done? Who is doing it? Why are they doing it? And how is it being done? We provide an in-depth analysis of all the possibilities that these genes may take the shape of within semantic web research. We use the 16 genes as the basis from the original genome and map them to existing semantic web research. An overview is presented in Table 3.3.

What is being done?

The first question that needs to be answered for any particular activity according to this genome model is: What is being done? The most obvious level of granularity it is the task which is to be performed. We consider the two genes proposed in the original genome:

Create: The actors in the system create a new artifact. It could be a piece of knowledge such as a concept, relation or classification.

Decide: This translates to an evaluation task or a selection which is to be performed based on certain alternatives, such as selecting the most appropriate relation from the set of given ones. The original genome model suggests that completion of a complete goal involves at least one of each of these create and decide genes. Create genes will generally be followed with a decide gene in order to make a decision about what items are to be kept. And the decide gene usually needs a create gene to generate the choices being considered.

Table 3.3: Overview of the Collective Intelligence Genome

Question	Gene	Manifestion
What	Create	Creation of new artifact
	Decide	Evaluation or Selection Task
Who	Crowd	Large group that performs a task
	Expert	Someone with specialized knowledge of the domain
	Hierarchy	Someone in authority
Why	Money	Financial gain as motivation for doing work
	Love	Enjoyment, fun, competition (from doing the task)
	Glory	Reputation and sense of community
How - Create	Collection	Independent tasks
	Contest	One entry in the collection considered best and rewarded
	Collaboration	Tasks are not independent (interdependent tasks), No way to divide an activity into smaller ones
How - Decide	Group Decision	Includes voting, averaging, consensus, prediction
	Individual Decision	Includes Markets and social networks

Who undertakes the activity?

The next critical question to be answered is: Who undertakes the activity?

Hierarchy or Management: Someone in authority makes a decision about who performs a task.

Crowd: Using this gene anyone who is part of a large group may take up a task who chooses to do so. This is a central gene in most collective intelligence systems.

Expert: This gene is not part of the original genome, however We have included this as an exclusive gene. We consider it to be different from the hierarchy or management gene because the distinction is important especially within the semantic web research. The semantic web research community recognizes the reliance on experts for certain domain specific tasks.

The choice between who performs when task is a critical one. The choice of crowds is usually to benefit from the skills and resources of a larger group of people at a much larger scale. Experts are usually hard to find. Crowds are usually suited for tasks that do not require any particular skills or domain specific knowledge. For the crowd gene to work, the task must be reasonably decomposable, and presented to the crowd in a feasible manner. At the same time, results must ensure feasible aggregability. In addition, mechanisms to prevent spam and sabotage are also indispensable. For those tasks where either the task is not decomposable enough or the reliability of the task performance becomes questionable, experts are usually relied upon. Traditionally in the semantic web research much of semantic content creation tasks have been performed by experts. Only recently the researchers have started experimenting with and reported findings about the tasks that are amenable to crowdsourcing and human computation.

Why the task is being performed?

The question of who is incomplete without the question of why? Why would someone perform a given task? What motivates them to participate? What incentives are provided to them? In a simplified sense the three broad level motivations are *Money, Love* and *Glory*. Siorpaes and Simperl proposed some [186] incentive based measures for sparking motivation for human contribution in the semantic web research. Their breakdown of incentive scheme is more or less encompassed in the *Money, Love* and *Glory* model.

Money: It is long known that financial gain is one of the primary motivations for most players in the markets and traditional organizations.

Love: This is manifested in various forms. Users enjoy engaging in certain tasks, especially feeling of community and socialization gives them a sense of belonging and purpose. This may include fun and competition incentives as suggested by Siorpaes and Simperl [186]. Reciprocity may also be said to fall into this category, whereby the contributors receive an immediate or long term benefit on performing a certain task.

Glory: This includes reputation and sense of community. Recognition and acknowledgement are primary contributors.

There may be several ways how the genes influence contributions. Love and glory when put into play as motivators are known to reduce costs but not always. However, money and glory together can bring results faster. The right combination especially within semantic web research remains to be experimented upon; however, this consideration is crucial.

How the task is being performed?

This is an important question that is a key determinant in driving collective intelligence in any system: How is it being done? A key deciding factor when ansering this question depends on the nature of decision making that follows crowd participation. Are the contributions and decisions independent? Or are there strong inter-dependencies between the contributions. The four types of How genes are: *Collection, Collaboration, Individual Decision*, and *Group Decision*.

For *Create* tasks, *Collection* and *Collaboration* are relevant whereas for *Decide* tasks *Individual* or *Group Decisions* are relevant.

Collection: The task performed is independent of each other. Such gene is only useful when the overall activity can be divided into tasks independent of each other. If this is not the case then collaboration is needed.

Contest: This is a subtype of collection gene. One of the entries is declared best. Or a participant with best performance may also be rewarded with a prize or some other recognition.

Collaboration: This occurs when members of a crowd work together and interdepencies exist between their work. This is usually needed when there is no way of dividing a large activity into smaller independent subsets.

Group Decision: This is useful when each participant has to make the same decision. Variants of this mode include voting, consensus, averaging and prediction.

Individual Decision: This is employed when wide-spread agreement is not needed. The decisions may not be needed for all. Variants of individual decisions include markets and aocial networks. Markets may not be relevant for semantic web research however, the domain of social semantic web has emerged as a result of social networks contributing towards semantic web research.

3.4.3 Application of Collective Intelligence Genome to Semantic Web Research

Table 3.4 shows the application of collective intelligence genome to selective works from the ones presented in Table 3.1 and 3.2. When compared with the genes of the original genome, to the actual mapping conducted on actual case studies from semantic web research, it is obvious that there is good spread of research available in both genres, i.e., GWAP and Mechanised Labor. The GWAP seems to dominate more. The mapping of collective intelligence is useful in providing us with some interesting insights. Although, the coverage of studies presented in Table 3.4 is in no way exhaustive, it gives reasonable insights into the current trends to date. While a range of semantic web tasks have been experimented with, the creation and the decide modes are fairly consistent between collection and consensus. This reflects that most tasks are of simple and independent nature. This leaves room for more studies to experiment and dwell further with more genes as presented in the genome.

One limitation that is felt for the collective intelligence when applied in this context to semantic web research and tasks especially when carried out using GWAP genre, oftentimes the task as presented to the user is different from what is intended or how the users responses will actually contribute towards the ultimate objective of semantic content creation. This is not reflected in a comprehensive sense, as of now, however, the genome may be revised to include this thread, specialized for semantic web research or for GWAP. Nevertheless, the threads of the collective intelligence genome provide useful means to cross analyze the research in the domain.

3.5 Crowdsourcing based Workflow Systems for the Semantic Web

In this section, we review and compare some of the most notable studies which adopt a crowdsourcing based model or mechanized labor for solving semantic web tasks. In particular, we include those studies in this analysis that adopt some form of workflow based approach in their design. We provide an overview of the different threads we use for comparative analysis and then provide a detailed analysis using the dimensions highlighted in these threads.

3.5.1 Threads for Comparative Analysis

The parameters used for the comparison are the key threads of analysis, which provide essential insights into the design and configuration of the key state of the art studies. The parameters are explained in Table 3.5 and the associated dimensions used later for analysis are also enlisted. Primarily, the key parameters we consider when comparing these systems are nature of task support, task design, workflow and data representation, worker interaction and engagement strategy, worker performance, data aggregation, and data quality and reliability.

3.5.2 Comparative Analysis using Ontology Engineering Tasks

Table 3.6 presents a comparative analysis of research studies primariy focusing on some core task in ontology engineering or linked data management using a microtask based crowdsourcing genre. We focus our analysis on five primary studies that include CrowdMap [164], Noy and Colleagues [148] & Mortensen et al.[140, 141, 142, 143, 144], ZenCrowd [51], CrowdLink [26, 27], and CrowdTruth[100]. We chose these systems since they present the state of the art when it comes to the confluence of semantic web and human computation research. The essential elements in the design of these systems provide the key ingredients for the design of similar applications in the near future.

CrowdMap[164] presents a workflow model for crowdsourcing mappings between ontologies. Works from Noy and Collegues also aim to solve ontology verification tasks using crowdsourcing. CrowdLink[26, 27] and ZenCrowd are similar efforts, that focus on linked data management tasks. However, much of the emphasis of these efforts lies in automating tasks for microtask crowdsourcing. There is considerable need for combining human computation with machine computation. CrowdTruth [100] present a framework for harnessing disagreement in gathering annotated data. They achieve a human computation workflow through a pipeline of four processes: 1) machine processing of media, 2) reusable task templates to collect human input, 3) application of disagreement metrics, and 4) result presentation. Similar workflow would be beneficial to engineering semantic web tasks.

We present a comparative analysis of crowdsourcing based workflow systems below using the threads and dimensions mentioned in Table 3.5.

Nature of Task Support

The nature of the semantic web tasks that have been undertaken are varied and many.

CrowdMap: CrowdMap primarily focuses on Ontology Alignment tasks and in that too, only class level or concept level alignments are undertaken. The instances and property level alignments are ignored.

Noy: The tasks undertaken in this study concerns primarily validation or verification tasks. This includes e.g. primarily a hierarchy verification i.e. to verify a superclass-subclass relationship.

ZenCrowd: The ZenCrowd focuses on extracting entities from HTML pages and linking them with appropriate entities on the LOD cloud. Candidate links are generated using automated extractors and some automated decision making is applied using probabilistic models to reduce the candidate mappings that need verification from the crowd. Therefore, primarily, in terms of crowdsourced tasks, only validation is obtained.

CrowdLink: The proposed CrowdLink architecture proposes to support a number of tasks in the linked data management from acquiring missing facts and information to verification and validation. Support for structural and schema level tasks is also mentioned. Experimentation is primarily carried out for LOD verification, knowledge acquisition and entity or link disambiguation.

CrowdTruth: CrowdTruth provides a wide range of annotation tasks for a range of input media such as text, multimedia and images. While the focus is on obtaining annotations and pure ontology engineering methodology is not incorporated in the framework, however, the crowdsourcing framework is generic and mature to be extended to incorporate an ontology engineering methodology.

Parameter	**CrowdMap [164]**	**Noy [148] Mortensen et al. [140, 141, 142, 143, 144]**	**ZenCrowd [51, 52]**	**CrowdLink [26, 27]**	**CrowdTruth [100]**
Task Category	Ontology Engineering	Ontology Engineering	Linked Data Management	Linked Data Management	Neither pure Ontology Engineering nor pure Linked Management
Tasks	Ontology Alignment	Ontology Verification	Entity Linking	Schema and Instance Level Tasks	(Semantic) Annotation (Support for Text, Image and Multi-media)
Task Granularity	Validation & Identification (Concepts only) No Properties or instances	Hierarchy Verification Verify a Superclass-subclass relationship	Verify Entity Links	Create New Knowledge Review Knowledge Entity Disambiguation	Text Annotation (Span and Relation Extraction) Event Extraction Video Annotation
Domain of Study	Conference Ontology (OAEI)	Biomedical Ontologies Comparison of Worker performance Upper vs. Application Ontology BWW, SUMO, WordNet, CARO	News Articles	Domain Independent Evaluations done for DBLP, GeoCities and US Census LOD datasets	Domain Independent Evaluations done for Medical Relation Extraction, Newspaper Event Extraction
Task Representation	Simple	NA	Static	Dynamic	Dynamic

Task Template	Simple (Limited to pairs generator)	NA	NA	Task Templates: • Entity Disambiguation • Create New Knowledge • Review Knowledge	Task Templates: • Factor Span Correction • Relation Identification • Relation and Direction Identification • Event Identification • Event Loc, Participants, and Time Identification
Reusable Task Template	Yes	NA	Yes	Yes	Yes
Use of Task Composition/Decomposition	NA	NA	NA	NA	Composite Tasks, made up of atomic tasks
Workflow Representation (Automated vs. Static)	Automated	NA	Static	Automated	Automated
Workflow Human and machine computation combined	Yes Candidate mappings are automatically generated Microtasks are automated	No Conceptual Workflow described in [144]	Yes	Yes Uses SPARQL Queries to generate input to the Micro-tasks by Querying LOD sources	Yes
Any particular data format	NA	NA	NA	SPARQL Queries as part of the input workflow Ontology Driven task profiles	Customizable Job Configuration and Settings API

Worker Engagement/Interaction Presenting Knowledge to Users Engage Workers	CrowdFlower 7 Alignment Questions into one HIT to facilitate worker assignment and resource optimization	Crowd (MTurk) Qualification Test (8 out of 12 to Qualify) $50 award to student with the best result Question formulation with positive and negative polarities Additional cognitive overload with negative polarity questions reduces performance	Crowd (MTurk)	Crowd (MTurk)	CrowdFlower Crowd (MTurk)
Worker Performance Benchmark	Workers Only	Turkers vs. Students Turkers vs. Domain Experts	Workers	Workers	Workers
Data Aggregation Measures	CrowdFlower Aggregation Precision and Recall measures for evaluation	Baysian Inference Model [141]	Agreement Vote Precision and Recall measures for evaluation	Majority worker Agreement Worker Confidence based measure Precision and Recall measures for evaluation	Sophisticated Disagreement Metrics Utilization of Annotation Vectors Worker and Unit Level Metrics Annotation Metrics
Data Visualization	NA	NA	NA	NA	Interface to support data and results Visualization

Data Quality and Reliability Measures, Spam Prevention	Golden unit question included in each HIT (of 7 questions each) for evaluating worker performance, detecting spam and validation	Redundancy in responses: 32 Random responses e.g. 23 identical answers out of 28 Disqualified 90% Approval rating from other requestors Making question answering time consuming by placing all 28 questions in the same page	Spam Detection using 3 consecutive random answers.	Simple Aggregation based on majority votes. No spam prevention Controlled test settings	Harnesses worker disagreement using Annotation Vectors and Media Unit Vectors Specialized Worker, Annotation and Unit Metrics Used

Table 3.6: Comparative Analysis of Crowdsourcing Research in the Semantic web

Task Design

CrowdMap: The CrowdMap system defines two types of micro-tasks namely: 1) Validation microtasks and 2) Identification microtasks. The validation microtasks presents a complete relationship between two classes and requires them to specify if they agree with the relationship or not. The identification task asks for the workers to identify a particular relationship between the source and the target class. The task includes the contextual information available for both classes.

Noy: The study primarily focuses on the verification task, specifically hierarchy verification is addressed. Therefore, there is only a single task used. Questions of the same type were placed on the same page for quality control purposes. Some experimental variations included question with or without term definitions as contextual information.

ZenCrowd: The study creates a single task for each entity that needs to be evaluated by the crowd. Some textual content along with the entity is included as means of contextual informations. The workers in turn have select the URIs that match the entity. Therefore the task at hand is primarily a selection task.

CrowdLink: CrowdLink provides a number of task templates e.g. new or missing triple information on the LOD may be sought from the crowd. Similarly, verification or validation tasks may be created using the available templates. A unique feature of the task design is that the input is dynamically obtained from the LOD sources using SPARQL Query and may be optionally attached with the task input parameters. Most task templates include a 'confidence level' question that asks the workers to provide their confidence level for their responses. This is done as means of quality control.

CrowdTruth: The CrowdTruth provides use cases that illustrate some 14 distinct annotation templates across three content modalities (text, image, video) and three domains (medical, news, culture). The framework design claims to provide its template collection as a continuously extendible library of annotation task templates, which can be reused and adapted for new data and use cases. The implementation of CrowdTruth does not pose restrictions for the creation of new templates. Some of the key tasks include Span Correction, Relation identification, Relation Direction Identification, Event and Type Identification, Event Location, Participants and Time Identification. In addition for images and videos, the annotation includes object identification or Event identification.

Some additional discussion points: It must be noted that most of these systems present support for atomic tasks. Only CrowdTruth has some level of support for composite tasks. Most tasks are simple and do not require any iteration.

Workflow and Data Representation

CrowdMap: The CrowdMap workflow architecture includes components for generating candidate pairs from a source ontology (Pairs Generator), a MicroTask Generator and a MicroTask publisher for publishing the tasks. Another end of the workflow deals with reading the results, processing, alignments and results evaluation. The functionality is meant to be generic to be integrated into existing environments of ontology alignment.

Noy: While [148] does not mention the use of any formal workflow model in their study, a conceptual workflow model is proposed by the parallel study of the authors in [140, 144]. The workflow includes processes for entity selection from a source ontology, task generation, optimization, spam and filtering and response aggregation. An additional process accounts for some form of context provisioning from external sources.

ZenCrowd: ZenCrowd leverages algorithmic and human workers working together to produce high quality results. The architecture of ZenCrowd takes as input HTML pages and enriches them using Entity Extractors, that extract textual entities, and linking them to the Linked Open Data cloud. For doing this it uses Algorithmic matchers for generating candidate mappings between the entities and the LOD links. It then sends these to a decision engine that sends selective candidates to a microtask manager, which creates tasks and publishes to the crowdsourcing platform. The decision engine evaluates the crowd responses using a probabilistic network and results are used to decide most appropriate links for the entities. The output are pages with entities linked to the respective URIs on the LOD cloud.

CrowdLink: The CrowdLink architecture is similar to CrowdMap and ZenCrowd with regards to task publishing and results retrieval. CrowdLink, uses reusable task profiles and task specifications for a wider range of tasks, and optionally uses SPARQL queries for the purpose of obtaining input from LOD sources. The architecture is equally capable of using any SPARQL compatible ontology source. The workflow management engine in CrowdLink architecture includes a task publishing workflow manager and a task review workflow manager, responsible for publishing and reviewing tasks respectively. A specific feature of the task review workflow is the update mechanism that directly updates the source ontology based on the evaluated results and specified acceptance thresholds. This is done using the SPARQL update protocol, and makes it more convenient to incorporate into ontology engineering methodology.

CrowdTruth: The CrowdTruth workflow framework is most sophisticated and extensive of all systems presented thus far. The CrowdTruth architecture utilizes a range of pre-processing techniques for processing a wide variety of input media such as text and multi-media collections. A number of customizable job templates are support for a range of tasks. The data post-processing applies a number of metrics for the analysis of the obtained results and extensive data analytics are applied. The very emphasis of the CrowdTruth methodology is to harness disagreements in the crowd responses. The CrowdTruth framework is perhaps the only one of ones presented that allows for composite tasks such that a complex tasks are broken into smaller ones.

Discussion Points: Most of these studies combine human and machine computation at some level. However, if we strictly restrict ourselves to to combine human and machine computation in cooperative and iterative workflows, such that human tasks are adaptively generated on the fly, then this is still something that is yet to be seen however, presents a promising direction for research.

Worker Interaction, Engagement Strategy and Performance

CrowdMap: CrowdMap uses the CrowdFlower system to engage workers. Researchers agree that worker responses can depend on the nature of the task design and optimal task design depending on the nature of the problem to be solved is always crucial. CrowdMap attempts the strategy of packing 7 Alignment questions into one HIT or job to facilitate worker assignment and resource optimization.

Noy: Noy and collegue experiment with a number of experimental setups to establish the effectiveness of crowdsourcing methods for ontology verification without compromising quality. They compare performance of turkers vs. students and also turkers vs. domain experts. They primarily rely on MTurk for engaging the workers. The specific research questions that the authors attempt to address are crucial in particular the aspect of determining the suitability of using crowdsourcing for verification of ontologies in specialized domains such as biomedicine. In addition, the authors also attempt to determine whether the workers performance change depending on how specific or generic the ontology is. Their results suggest that crowdsourcing is indeed feasible for specialized domains.

ZenCrowd: MTurk is used as the platform for engaging the workers.

CrowdLink: MTurk is used as the platform for engaging the workers. Only sandbox testing is conducted. No payments are used.

CrowdTruth: CrowdTruth uses the CrowdFlower and the MTurk platforms for engaging the workers.

Data Aggregation, Quality and Reliability

CrowdMap:CrowdMap relies on CrowdFlower aggregation methods and uses Precision and Recall measures for evaluating the results.

Noy: Primary aggregation measure used is a majority consensus. In one version of the experiments, the author utilize a Bayesian inference model for verification [141]. A number of experimental settings are tried by the authors for quality control. Some of the control parameters include, qualification questions, question design, redundancy and disqualification, and spam identification. Qualification questions were employed however according to [141] these did not affect the results much. They introduce redundancy in responses. They also disqualify the responses of the worker if 23 out of 28 answers are identical. In addition, 90% approval rating is also enforced. The question answering process is also created to be time consuming to ensure only serious participants record their responses. These are some of the ways that spam has been attempted to be prevented. However, as authors note these tests alone are not sufficient.

ZenCrowd: The Agreement voting technique is employed in this study. 5

different workers are required to select from amongst the proposed mappings and the URIs with at least 2 votes are selected as valid links. A technique of blacklisting bad workers (or spammer) is used. If a worker randomly and rapidly selects the links, its considered noise in the system and the worker is blacklisted. Consecutive 3 bad answers in the training phase are considered enough to identify the worker as a spammer.

CrowdLink: The experiments are done under controlled settings, therefore the spam element is not taken into consideration. The task design includes a confidence measure which is obtained from the workers to indicate the worker confidence into consideration. Simple agreement based aggregation is done.

CrowdTruth: CrowdTruth has an extensive Data Analytics component. There are several annotation (disagreement) metrics used [14, 15, 188]. The CrowdTruth Framework aggregates the annotations of multiple workers across different media units (e.g. text, image and video) using *Annotation Vectors*, and *MediaUnit Vectors*. The Annotation vector records the responses of the workers and the similarity is computed using cosine similarity measure. The MediaUnit vector accounts for all worker submissions on a unit for a given task. Analysis of ambiguity to prevent spam is done across each system component i.e. worker, annotation and unit, resulting in 'Worker Metrics', 'Annotation Metrics' and 'Unit Metrics'.

Worker metrics determine the measure of disagreement at the level of the worker in order to distinguish spam from a high quality response. Various measures such as worker-unit disagreement, worker-worker disagreement, and average annotations per unit are used. The unit metrics are used to determine the clarity of the input unit given to the crowd. To identify ambiguous units, measures such as unit annotation score, and unit clarity are used. The Annotation metrics aim to measure the quality of pre-defined annotation types, in attempt to distinguish between disagreement resulting from either low quality workers or disagreement resulting from badly designed task. The specific measures used in this category are annotation similarity, annotation ambiguity, annotation clarity and annotation frequency.

Together these metrics provide an in-depth perspective to improve the crowdsourcing task design and analysis of results. The CrowdTruth is the only framework which also makes use of Data Visualization of the annotation metrics once aggregated.

3.5.3 Discussion

As illustrated in the Table 3.6, the granularity of tasks that has been subject to crowdsourcing in the recent studies are fairly at low level of granularity. There is little evidence of task composition or decomposition, since most tasks are fairly simple. Either simple task representations have been used or there is no evidence of any specific task representation being devised. Similarly, it is not evident if any well defined task templates have been developed for ontology engineering tasks. Another significant limitation that can be seen is the lack of formal workflow engineering. The workflows so far have minimal formal representation and are simplistic in nature. There is considerable need for more efforts directed in this direction.

3.6 Application of Semantic Web Techniques to Facilitate Human Computation

Research towards this direction, whereby there is evidence of semantic web techniques applied to improve the state of human computation systems are fewer. Most notable work in this domain is in the direction of generating human-computer micro-task workflows using domain ontologies Luz et al. [125]. There is considerable interest in resolution of complex tasks that require the cooperation of human and machine participants has emerged. The approach proposed by Luz et al. [125], consists of a semi-automatic generation environment for human-computer micro-task workflows from domain ontologies. The inherent process relies in the domain expertise of the requester to supervise the automatic interpretation of the domain ontology. According to the authors, the unstructured nature of micro-tasks in terms of domain representation makes it difficult (i) for task requesters not familiar with the crowdsourcing platform to build complex micro-task workflows and (ii) to include machine workers in the workflow execution process. As claimed by the authors, the structure and semantics of the domain ontology provides a common ground for understanding and enhances human-computer cooperation. The study provides an interesting dimension about the use of ontologies for the purpose of workflow generation and may inspire future relevant studies.

One evidence of utilization of ontologies for facilitating human computation systems is for provenance management. The vocabulary SC-PROV, a provenance vocabulary for social computation [135] is an important step in managing the provenance information and thus contributing to enhancing the transparency in human computation systems. This is crucial for enabling decision making about the reliability of participants and quality of the generated solutions. Baillie et al., [20], while attempting to solve the issue of quality reasoning for the semantic web also highlight the importance of obtaining provenance information in case of crowdsourced knowledge followed by enhancement measures to ensure quality.

3.7 Discussion and Analysis

In chapter we set out to examine the application of human computation methods and crowdsourcing to the domain of semantic web. In particular, we first analyzed and presented the need for human contribution in the semantic content creation. We dealt with this in two specific categories i.e. ontology engineering and linked open data management. We also examined the key human computation genres that have been applied to the semantic web tasks. We also provided an in-depth classification of these tasks according the the examined genres. We also provided brief insights into the automated tool support for human computation tasks in the semantic web. Another key contribution of the chapter was the collective intelligence genome applied to the semantic web case studies. The genome helps analyzing the essential threads of composition that require thought and effort in conceiving a human computation driven application. We also carried out a dedicated discussion focusing on those state of the art systems that have attempted to design and implement workflow systems for driving the semantic web tasks using the human computation methods, since we strongly feel, the future will see more of such systems, with more mature and well grounded design principles. Before concluding the chapter, we highlight some of the key outstanding challenges which present areas for further research.

3.7.1 Key Outstanding Challenges

Different genres of human computation present different challenges. We briefly discuss the challenges faced by semantic GWAPs and focus more on the challenges faced in the domain of mechanized labor or human computation systems in general.

Challenges in Semantic GWAPs

Siorpaes and Hepp [183] provide some key challenges that are faced when creating semantic GWAPs, including, task identification in the semantic content creation, designing game scenarios, designing a usable and attractive interface, identifying reusable bodies of knowledge, preventing cheating, avoiding pitfalls, e.g. unintentional agreement on wrong choices, fostering user participation, deriving semantic data, efficient distribution of labor and scalability and performance to name a few. Some of these overlap with the other genres in human computation, however, semantic web games have their own particular challenges such as game design, game scenarios and user engagement.

Methodological Considerations

While on one hand the domain of crowdsourcing has matured over the years, however, its intersection with ontology engineering has yet to see a formal methodology taking shape. [5] highlight the need for methodology and best practices, while moving away from isolated approaches.

Integrated Tool Support for Human Computation Workflows

It has to be recognized that a significant limitation that can be felt is integrated tool support for semantic web specific human computation and crowdsourcing. Recent work by Hanika, Wohlgenannt and collegues have addressed this limitation [5, 80, 81] by providing a plugin for the popular ontology development tool Protégé. This is a crucial advancement towards measures that aim to integrate crowdsourcing into ontology engineering practices. However, this is only the beginning and there is much room for further research including human computation workflows in ontology engineering, quality of crowdsourcing results and the large scale application and usability of such plugins. The results of these tools may vary greatly between different tasks (depending on the type of task and the difficulty of the domain). They also show sensitivity to the timing when the task is crowdsourced and the responses obtained from the available workers [5].

Most applications themselves are single purpose and standalone. The further success demands integrated tool support in particular with existing semantic web tools to realize the complete potential of crowdsourcing in particular and human computation in general.

Kondreddi et al, [112] have proposed methods to combine information extraction approaches with human computation for knowledge acquisition, and claim to reduce to cost of human computation effort. Such efforts are claimed to be 'game changers' in the new generation of systems that combine power of humans and machines.

Aggregating Answers

One of the key issues faced when inputs from a large crowd is obtained is that of aggregating answers in a quality manner. Lopez et al., [124] argue that merging and ranking of answers in the semantic web domain that are obtained from different sources, similar to wisdom of the crowds tend to produce higher precision. There is need for more well defined quality measures especially in specialized domains.

Adequately Utilizing Crowdsourced Knowledge and Annotations

Often times it becomes a challenge in itself to utilize the crowdsourced knowledge especially in those tasks where domain specific annotations are needed. Thus, workflows and techniques emphasizing such adequate utilization need to be thought of. As an example, 3DSA [213] is a system that utilizes rule based reasoning to classify crowdsourced annotations.

Design of (Optimal) Semantically Engineered Workflows

Whereas a combination of human and computational intelligence is often likely to yield superior results [177] , the design of an optimal workflow depends on various dimensions, such as the type of the task and the degree to which it can be automated, the amount of data for which the task is expected to be (repeatedly) executed, and the (estimated) availability of capable workforce potentially interested in en-gaging with the task [108]. Luz et. al [126] highlights the need for a structured and semantically enriched representation of the micro-task workflow and data to allow for a better integration of human computation and machine computation efforts. The use of ontologies as means of representing task profiles is also suggested by [26, 27]. One such effort of representing micro-tasks using ontologies and their dependencies through the specification of their domain (input, output and context) has been presented by [125]. More efforts to this end are expected to emerge. Combining human and algorithmic computations needs more work.

Motivating Users and the Role of Incentives

It has yet to be seen what type of incentives, platform, games, rules, systems and architectures will work best for human computation on the semantic web [183]. In order to benefit from network effects and positive externalities, end-user semantic content authoring technology needs to offer an appropriate set of incentives in order to stimulate and reward users participation, generating in massive production of useful semantic content [185] .

Dealing with Diversity

Dealing with Motivational, Cognitive and Error Diversity: Because people are involved, programming the 'Global Brain' is deeply different from programming traditional computers [34].

Harnessing Crowds vs. Experts: Choosing the Right Crowd

Crowdsourcing simple (independent) vs. complex (interdependent tasks) presents a challenge especially to semantic web researchers. Research evidence suggests that there is strong correlation between the nature of task, its difficulty, skill and knowledge required to the accuracy obtained from crowdworkers [149, 150]. Although, crowdsourcing is promising, however for domain specific tasks, it needs to be establish which task is suitable for which kind of crowd. The idea of nichesourcing [49] or expert sourcing [5] and expert finding [40] is being proposed for knowledge intensive tasks. These ideas bring their own challenges such as adequate task distribution and quality assurance. Some efforts have attempted to tackle skill based task matching or task recommendation such as [71], [217] and [218] may serve to provide useful insights in this regard.

Compromise between Quality and Complexity of Tasks

Simperl and Wolger et al., [181] highlight that the area of human computation in semantic web engineering and especially linked data management is in its early stages and more research remains to be done in order to achieve the overall vision. They highlight the constrains that result in the complexity of the tasks that can be feasibly undertaken and the domains for which knowledge can be reliably collected from a mass audience, when a mass userbase is targeted against tasks and tools designed for experts. This introduces the need for specific evaluation and quality assurance mechanisms.

Want et al., [206] also highlight the importance of Data complexity and specificity of the tasks in appropriate task design and management.

Quality Measurement

Quality management for crowdsourcing is critical. Techniques for accurately estimating the quality of workers is pivotal for the success of crowdsourcing research [103]. Developing common grounds for quality assessment is crucial when obtaining semantic annotations from the crowd [101] and result in annotator diversity that can be harnessed. Aroyo and Welty present a series of works in order to harness disagreement in crowdsourcing gold standards [13, 14, 15].

Constraints to Microtask Approach

Decomposability, verifiability and expertise are three crucial factors that constrain the microtask design [179]. Techniques that resolve to achieve a suitable compromise are need to see the research through a state of maturity. Several parameters can have an impact the on the approach used and the resulting quality such as effect of context on worker performance, effect of more domain specific tasks on the accuracy of the crowdsourcing model, improving worker questions by fine tuning the questions, and optimizing the number of questions issued to the user [164].

Scalability

Scalability in both reading and writing semantic data. It is yet to be seen what a truly web scale semantic human computation system will look like [54].

3.7.2 Open Research Issues for Semantics Driven Human Computation Systems

While the view of challenges presented in the preceding section itself sheds light on the current and open research directions towards incorporating human computation methods into semantic web processes, on a parallel note, we also enlist a few directions to be possibly taken up for further research, towards augmenting semantic content to better design and facilitate human computation systems:

Design of New Vocabularies

New ontologies and vocabularies will need to be developed to help manage and link these human computation systems together[54].

Semantic User Management

Users can easily sign on into new systems, and have their points and reputation follow them [54].

Meta-data and Provenance Management

The idea of trust, reliability and ensuring quality of crowdsourced contributions demand that provenance information be essentially embedded within workflows and responses and be included for relevant analysis[54].

Recent proposition by Markovic and collegues highlight the prevenance perspective [131, 132, 133, 134, 135] when dealing with the crowd and linked data. Provenance information is critical to be maintained and therefore adequate measures for managing provenance becomes a key challenge.

New Programming Models and Frameworks

Similar to the ideas of programming the 'Global Brain', programming the 'Global Brain Semantic Web' [34] demands the need for programming languages and frameworks such as the one presented in [138]. Exploring iterative and parallel human computation processes [122] is another prospective idea for further research.

Design of Better User Interfaces

New visualizations and usable design is needed for better user engagement.

State Space in Human Computation Systems

According to Difranzo and Hendler [54], 'state space' in a human computation system is the collection of knowledge, artifacts and skills of both the human users and the computer system that help define the state, or current stage at a given time, of the human computation system. This state space can often be very messy, disjointed, incomplete or inconsistent. The semantic web could provide this common platform and medium for representing knowledge that persists despite the asynchronous behaviors of the human participants. More research is needed to explore how this could work, how best to represent this knowledge, and what advantages this could bring to future human computation systems.

3.7.3 Some Reflections

Most of the current semantic web research has primarily focused on studying the feasibility and amenability of applying crowdsourcing and human computation techniques to the semantic web tasks. There is much work that remains to be done. Carletti et al., [43] highlight an essential challenge about the separation that occurs between crowdsourced activities and organizational workflows. Although, the discussion is in the context of digital humanities, the semantic web landscape appears to be facing a similar challenge. While there has been significant interest and contribution on these fronts, it still remains to be established how much of the crowd-sourced contributions have really made a real impact. Crowdsourcing has also shown to contribute to innovation [6]. However the key consideration for crowdsourcing has to be the goals for the crowdsourcing activity [6]. The interleaving of human, machine, and semantics even have the potential to overcome some of the issues currently surrounding Big Data [146].

3.8 Conclusions

In this chapter we presented a review of the challenges that the semantic web domain has faced especially in terms of the need for human intervention, and in this light analyzed the intersection of semantic web and the human computation paradigm. Two fold approach towards understanding this intersection provided interesting insights. Both semantic web and human computation have potential to benefit from each other. While there seems to be more interest by the semantic web community to benefit from the human computation paradigm, there is evidence of the latter being enhanced using the former as well. The comprehensive view of the promises and challenges offered by the successful synergy of semantic web and human computation as discussed in this chapter indicates that this is a thriving research direction and is expected to grow further. We hope that this review will serve as basis for exploring newer threads of synergy between the semantic web and human computation research, resulting in the creation of better applications and approaches that advance both domains.

Table 3.4: Application of Collective Intelligence Genome to Semantic Web Research

Source	What	Create/Decide	Who	Why	How? C-Create D-Decide
GuessWhat [130]	Domain Ontology Creation from Linked Data	Evaluate if the class name fits the description	GWAP	Love/Glory	C-Collection D-Consensus D-Majority Voting
Verbosity [203]	Not pure Ontology Engineering or Semantic Web GWAP	Select an appropriate relation from the set of given relation types	GWAP	Love/Glory	C-Collection D-Consensus
OntoPronto [184]	Conceptual Modelling, Concept and Instance Collection	Decide if given entity forms a class or an instance	GWAP	Love/Glory	C-Collection D-Consensus
Virtual Pet and Rapport [116]	Concept Collection	Concept collection	GWAP	Love/Glory	C-Collaboration
SpotTheLink [191, 192]	Ontology Alignment, Entity Linking	Select from set of terms, specify relation	GWAP	Love/Glory	C-Collection D-Consensus
VeriLinks [117]	Entity Linking	Validation of linked data	Crowd (MTurk)	Money	C-Collection D-Consensus
UrbanMatch [44, 45, 133]	Alignment and Interlinking	Link photographs with LOD concepts	Mobile Crowd	Love/Glory	C-Collection D-Consensus
CrowdSourced InPhO [62]	Conceptual Modelling: Conceptual Hierarchy Formulation	Select relations , select from term pairs	Crowd (MTurk)	Money	C-Collection D-Consensus
Noy [148] Mortensen et al., [140, 141, 142, 143, 144]	Ontology Verification	Hierarchy verification	Crowd (MTurk)	Money	C-Collection D-Consensus
CrowdMap[164]	Ontology Alignment	Mappings between Ontologies Equivalence Subsumptions Validation: Validate a given mapping Identification: Select between different types of given relations	Crowd (CrowdFlower)	Money	C-Collection D-Consensus
Conference v2.0 [46]	Ontology Alignment (Benchmarks)	Validation: Given two concept descriptions, validate if they are similar	Crowd (MTurk)	Money	C-Collection D-Consensus
CrowdSPARQL [4, 180]	Coneptual Modelling Ontology Classification Entity Resolution	Use SPARQL Query processing tasks as means of crowdsourcing	Crowd	Money	C-Collection D-NA
ZenCrowd [51, 52]	Entity Linking	Picking links in order to match similar entities	Crowd (MTurk)	Money	C-Collection D-Prediction
CrowdLink [26, 27]	Linked Data Management Tasks	Missing LOD Knowledge Validate LOD Links	Crowd (MTurk)	Money/Glory	C-Collection D-Consensus
Dr. Detective[61]	Extraction of Text Annotation	Medical text annotation (to obtain groud Truth)	GWAP	Love/Glory	C-Collection D-Consensus
CrowdTruth [100]	Semantic Annotation Tasks	Text annotation, Image annotation Video annotation	Crowd-Flower, Crowd (MTurk)	Money	C-Collection D-CrowdTruth Metrics
Acosta et al. [3]	Quality Assessment of Linked Data	Incorrect or incomplete object value Incorrect data and links	Contest (TCM Tool) & Crowd (MTurk)	Money Love Glory	C-Collection D-Consensus
uComp [5]	Ontology Engineering Tasks	A whole range of ontology engineering tasks	Crowd (MTurk) via Protege	Money	C-Collection D-Consensus, Authority

Table 3.5: Threads and Dimensions of Analysis

Parameter	**Dimensions of Analysis**
Nature of Task Support	a. What specific Semantic Web based tasks have been experimented with
Task Design (Task Composition, Representation and Reusability)	a. What methods of task representation are used b. What methods of task decomposition are used c. What kind of task templates and UIs have been used d. The extent of reusability of task design and templates
Workflow and Data Representation	a. How human and machine computation is combined into a workflow b. What are common data formats in the semantic web research when applied to human computation
Worker Interaction & Engagement Strategy	a. What strategies have been used for presenting semantic web knowledge to users in an easy manner b. What strategies have been used to engage workers c. How these strategies may be classified d. Which have been used in semantic web research and which have been not
Worker Performance	a. How workers have performed b. What are the factors the worker performance depends on c. What performance measures are used
Data Aggregation	a. How the data and results are aggregated b. What methods are used for aggregation of data c. What challenges and prospects in this direction d. How these methods may be classified
Data Quality and Reliability	a. How data quality and reliability is ensured b. What methods are used c. What are the challenges faced so far d. How the data quality and reliability measures may be classified

Chapter 4

Text Mining Verse Similarity for Multi-lingual Representations of the Qur'an

ٱللَّهُ نَزَّلَ أَحۡسَنَ ٱلۡحَدِيثِ كِتَٰبٗا مُّتَشَٰبِهٗا مَّثَانِيَ تَقۡشَعِرُّ مِنۡهُ جُلُودُ ٱلَّذِينَ
يَخۡشَوۡنَ رَبَّهُمۡ ثُمَّ تَلِينُ جُلُودُهُمۡ وَقُلُوبُهُمۡ إِلَىٰ ذِكۡرِ ٱللَّهِۚ ذَٰلِكَ
هُدَى ٱللَّهِ يَهۡدِي بِهِۦ مَن يَشَآءُۚ وَمَن يُضۡلِلِ ٱللَّهُ فَمَا لَهُۥ مِنۡ هَادٍ

Allah has sent down the best statement: a consistent Book wherein is reiteration. The skins shiver therefrom of those who fear their Lord; then their skins and their hearts relax at the remembrance of Allah. That is the guidance of Allah by which He guides whom He wills. And one whom Allah leaves astray - for him there is no guide.

— *[Al-Quran, Az-Zumar,39:23]*

Chapter Overview[1]

In this chapter, we explore the problem of text similarity in the context of multi-lingual representations of the Qur'an. Particularly, we use Arabic and English datasets of the Qur'an for comparative study and analysis of several similarity measures applied across different representations of the verses in the Qur'an. We provide useful insights into the impact of using different similarity measures applied to different features across different representations and linguistic characteristics of similar text. The ultimate purpose of the similarity computation framework developed is to be utilized for creating candidate links between verses or other texts to be fed to the crowdsourcing workflows devised and presented in later chapters of this dissertation.

4.1 Introduction

The Qur'an, considered as a concise data set, consists of less than 80,000 words, sequenced in 114 chapters (Surahs) and 6,236 verses (Ayahs) [17]. The original data format was spoken Classical Arabic. We treat the problem of computing similarity between the verses of the Qur'an as a special case of computing document similarity, a widely studied subject in literature. Document similarity measures are often utilized for the purpose of automatic text classifcation, and clustering [8] [99]. We treat a verse somewhat similar to a document. However, some verses may

[1]This chapter is published as:
Basharat, A., D. Yasdansepas, and K. Rasheed. "Comparative Study of Verse Similarity for Multi-lingual Representations of the Qur'an." *Proceedings on the International Conference on Artificial Intelligence (ICAI). 2015.*

be as short as a single or few words long. The longest verse in the Qur'an does not span more than a single page in the standard manuscript writing. Researchers recognize that the determination of a pair of documents being similar or different is not always clear and is often context dependent [99]. The problem at hand bears resemblance to the context of short text classification and clustering [189] [197].

We are not only interested in analyzing verse similarity using the original Arabic script, but also take considerable interest in undertaking this similarity study within and across different languages. Qur'an is one of the most widely translated texts, translated into numerous languages. A recent open linked dataset called 'Semantic Quran' has been published [168], with more than 48 known translations obtained from Tanzil[2]. Research in other domains has revealed interesting findings when cross-language similarity studies were undertaken [67].

The problem of computing the similarity between the verses has not been widely studied, despite the fact, that the Qur'anic text has attracted much attention in recent years from the computational research, artificial intelligence and natural language processing researchers in particular. There have been studies conducted using various corpuses in Arabic, some of which are [70], [107]. Sharaf et al.[167] are among the pioneers to have conducted some similarity and relatedness studies on the the Qur'anic text. The QurSim corpus is a corpus that is marked with relatedness information judged by a human domain expert. The authors highlight the challenges the variation in the degree of relatedness between texts. While there are instances where lexical matching is evident between the terms, the authors

[2]http://tanzil.net/

believe that the majority of the related pairs require a deeper and more semantic analysis and domain specific world knowledge in order to relate the two texts in the pair. For a more detailed discussion on the relatedness and the extent of its broadness, we refer the reader to the original work of Sharaf and his collegues.

Our objective here is to use the Qur'an's text to compare and contrast the different classes of similarity that may result using different representations of the text, and also using different similarity measures. We believe that a benchmarking study towards standardizing similarity measures for the Qur'anic text would pave the way towards the vision of achieving the ongoing efforts of standardizing the Qur'anic knowledge map as described by Atwell and collegues [17, 18]. We intend to use this study as the basis of developing semantic networks of similarity and relatedness between not only verses of the Qur'an but also extending to other contextually relevant texts, which are considered indispensable when it comes to developing a comprehensive and coherent understanding of the Qur'anic verses.

In this chapter, we evaluate a variety of document similarity measures, using multi-lingual, heterogenous representations of the Qur'an. In order to do this, we develop a similarity computation framework for the verses in the Qur'an (Section 4.2). We experiment with four different dataset representations, four similarity measures and three different feature representations and provide comparative insights into the results obtained (Section 4.3 and 4.4). We also shed light towards how we plan to extend this study further (Section 4.5).

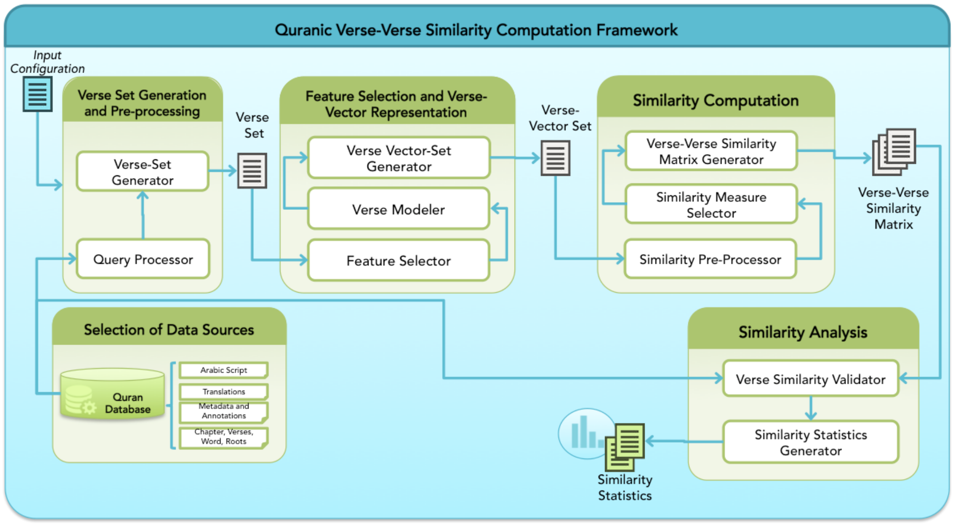


Figure 4.1: Qur'anic Verse Similarity Computation framework

4.2 Verse-to-Verse Similarity Computation Framework for the Qur'an

In this section, we present the design of our developed framework, which is meant to facilitate the process of similarity evaluation between the verses of the Qur'an, from the available datasets. The framework we have developed is generic enough to be readily used with other texts.

4.2.1 Datasets and their Characteristics

We have created a custom database for the experimentation which consists of the Qur'anic text in its original Arabic script in two forms: Firstly, with diacritics, i.e.

case markings, and secondly, without diacritics, i.e. plain, simple and clean arabic script. In addition, the database also contains translations in several languages. We used the translation of the Qur'an to the English language (translation by Yusuf Ali) for this study. The last type of data representation we used was a dataset generated for the Arabic roots (stems). A summary of the datasets used is shown in Table 4.1.

Table 4.1: Dataset representations and samples

Abbreviation	Description	Example
Q-DIAC	Arabic text with diacritics (or case markings)	وَ لَمْ أَكُن بِدُعَائِكَ رَبِ شَقِيًّا
Q-NODIAC	Arabic text without diacritics	ولم أكن بدعائك رب شقيا
Q-ROOTS	Arabic roots dataset	كون، دعو، ربب ، شقو
Q-ENG	English Text	"...and never have I been in my supplication to You, my Lord, unhappy."

4.2.2 Preprocessing of Verses

For this purpose, we used the raw data for each dataset without any linguistic pre-processing, except tokenization. We created a workflow mechanism such that we specify an input configuration that encodes the input data representation, the feature selection method to be used, and the similarity method to be used for an experiment. The Verse-Set Generator as shown in Figure 4.1, uses the Query

Processor to generate the dataset for experimentation from the database, and prepares the appropriate input representation to be used in later stages.

4.2.3 Feature Selection and Verse Representation

In order to adequately compare and contrast the different representations and the multi-lingual texts, we adopt a vector space model as means of verse representation for our experiments. For preliminary analysis we did not use any specific pre-processing techniques such as stemming. We relied on the traditional term-vector approach, which is the conventional approach adopted in most text mining applications.

We consider $V = \{v_1, v_2, v_3 ... v_n\}$ to denote the set of Verses in the Qur'an, and $\{T = t_1, t_2, ...t_m\}$ as the set of distinct terms occurring in V. We represent each verse, v, in the Qur'an data set as an m-dimensional vector $\bar{v} = \{a_1, a_2, a_3....a_m\}$, where a_i is the weight of the ith feature in the vector $\bar{v}$.

The framework we developed primarily operates on a term-verse matrix ($tv - matrix$), which is generated using the Verse Vector Set generator. It is a compilation of all the verse vectors into one matrix where the columns represent the terms (the vocabulary of the dataset) and the rows represent the verses. First the vocabulary is compiled from the verses, and the frequency of terms across all verses is recorded. Then the $tv - matrix$ is developed based on the verse order, in which the elements of the matrix are calculated according to one of the three term weighting methods namely: Boolean(B), Frequency(F) or TF-IDF(TF).

In traditional document similarity studies the $tf - idf$ method of term weighting is most recommended. However, Qur'an is considered as a text with several unique characteristics. Arabic linguists consider it as a profound piece of text, where each letter and word is relevant. Therefore we choose to experiment with different term-weighting measures to analyze the impact on the similarity computation. The feature selector is responsible for applying the feature selection method specified in the input configuration. The verse modeler then applies this selection to each verse before passing it to the verse vector set generator, which generates the $tv-matrix$.

We start with the most simple measure i.e. the Boolean. Using this measure the weight of the term t_i, would have the value 1 if the term is present in the verse and 0 otherwise. Equation (4.1) shows the term weighting using the term-frequency.

$$a_i = tf(v, t_i) \tag{4.1}$$

Equation (4.2) shows the term weighting using the term-frequency, inverse document frequency approach. In this scheme, N_v is the total number of verses in the dataset, vf_j is the verse frequency, and tf_{ij} is the number of occurences of the feature a_i in the verse vf_j.

$$a_i = tf_{ij} \times \log(N_v/vf_j) \tag{4.2}$$

4.2.4 Similarity Computation

With the verses represented as vectors, we measure the degree of similarity between the two verses as the correlation between their corresponding vectors using the Similarity computation module. The measures described below reflect the degree of closeness or separation of the target verses. Choosing an appropriate measure of similarity is crucial. We believe that understanding the effects of different similarity measures when applied to the Qur'an to be of great importance in helping to choose the best one. We adopt four of the measures described in [190] and [99] and map them to the problem of computing similarity for the Qur'anic verse vectors. The output of this stage is a Verse-Verse Similarity matrix, which provides a similarity measure on a scale of 0 to 1 for each verse pair in the Qur'an.

Euclidean Distance

Given the two verses v_a and v_b, represented by their term vectors $\bar{v}_a$ and $\bar{v}_b$ respectively, the Euclidean distance of the two verses is defined in Equation (4.3).

$$D_E(\bar{v}_a, \bar{v}_b) = (\sum_{t=1}^{m} |w_{t,a} - w_{t,b})|^2)^{1/2} \tag{4.3}$$

where the term set is $T = \{t_1, t_2 ... t_m\}$. As mentioned earlier, we use different weighting measures. Therefore, the $w_{t,a}$ may be $tfidf(d_a, t)$ or $tf(d_a, t)$ or $tb(d_a, t)$.

Cosine Similarity

Cosine similarity is one of the most popular similarity measure applied to text documents. Some notable studies that report the use of this distance measure include [99] [190]. Given the two verses v_a and v_b, represented by their term vectors $\bar{v}_a$ and $\bar{v}_b$ respectively, their cosine similarity is given by Equation (4.4)

$$S_C(\bar{v}_a, \bar{v}_b) = \frac{\bar{v}_a \cdot \bar{v}_b}{|\bar{v}_a| \times |\bar{v}_b|} \tag{4.4}$$

where v_a and v_b are m-dimensional vectors over the termset $T = \{t_1, t_2 ... t_m\}$. Each dimension represents a term with its weight in the verse, which is non-negative. Therefore, the cosine similarity is non-negatively bounded between 0 and 1.

Jaccard Similarity

The Jaccard coefficient measures similarity as the intersection divided by the union of the entities. For the verses, the Jaccard coefficient computes the ratio between the dot product and the sum of the squared norms minus the dot product of the given verse vectors. The definition is given in Equation (4.5).

$$S_J(\bar{v}_a, \bar{v}_b) = \frac{\bar{v}_a \cdot \bar{v}_b}{|\bar{v}_a|^2 + |\bar{v}_b|^2 - \bar{v}_a \cdot \bar{v}_b} \tag{4.5}$$

The Jaccard coefficient is a similarity measure and ranges between 0 and 1. It is 1 when the $v_a = v_b$ and 0 when v_a and v_b are disjoint, where 1 means the verses are the same and 0 means the verses are completely different.

Pearson Correlation Coefficient

Pearson's correlation coefficient is another measure of the extent to which two vectors are related. There are different forms of this coefficient. Given the term set $T = \{t_1, t_2 ... t_m\}$, a commonly used form is given in Equation (4.6).

$$S_P(\bar{v}_a, \bar{v}_b) = \frac{m \sum_{t=1}^{m} w_{t,a} \times w_{t,b} - TF_a \times TF_b}{\sqrt{(m \sum_{t=1}^{m} w_{t,a}^2 - TF_a^2)(m \sum_{t=1}^{m} w_{t,b}^2 - TF_b^2)}} \tag{4.6}$$

where $TF_a = \sum_{t=1}^{m} w_{t,a}$ and $TF_b = \sum_{t=1}^{m} w_{t,b}$.

This is also a similarity measure. When $v_a = v_b$, the value will be 1.

4.2.5 Similarity Analysis

We devised three similarity classes for the purpose of analysis as shown in Table 4.2. We automate some validation measures for determining the results. We use the Verse-Verse Similarity matrix for computing the statistics on similarity reported in Section 4.3.

Table 4.2: Similarity Classes for Analysis

Class	Description	Similarity Range
Identical	Identical Verses	Verse Pairs with Similarity Value 1
High	Almost Identical, Near Identical	Verse Pairs with Similarity Value > 0.9
Medium-High	With identifiable similarity	Verse Pairs with Similarity Value between 0.75 and 0.9
Low	Minimal semantic similarity	Verse Pairs with Similarity less than 0.75

4.3 Experimentation and Results

4.3.1 Evaluation Measures

We devised the accuracy measure, for the sake of quantification, for the *Identical* similarity class using the evaluation measures shown in equations (4.7 - 4.9).

$$P(Precision) = \frac{TP}{TP + FP} \tag{4.7}$$

$$R(Recall) = \frac{TP}{TP + FN} \tag{4.8}$$

$$F1(F - Measure) = \frac{2 \times P \times R}{P + R} \tag{4.9}$$

For the sake of analysis, we limit our quantification to those verse pairs with similarity value 1. Therefore, we treat those verse pairs with similarity value 1 as true positives (TP). A TN is therefore a verse pair with similarity value that does not equal one. The reason for limiting this analysis is that establishing the precise similarity between two verses has not been done in a standardized manner and there are no existing benchmark datasets available. The only dataset available QurSim [167] employs a different approach, i.e. it classifies verse pairs according to their relatedness measure, using human subjective judgement, based on the verse pairs compiled from a famous source of Qur'anic Exegesis by Ibn Kathir. For our study, we rely on pure measures of distance, as described earlier in this section. Therefore, establishing a similarity benchmark is not possible against which evaluation may be carried out. The only pairs of verses we are able to consider as ground truth are the identical verse pairs.

4.3.2 Experiments

The experiments are applied to every combination of using each of the four dataset representations, each of the similarity measures and each of the term-weighting methods. We therefore had $3 \times 4 \times 4 = 48$ experiments in total. An abbreviation scheme is used to denote the experiments. A sample experiment scheme indicates the dataset used, term-weighting applied and the distance measure. E.g. Q-DIAC-C-B indicates, Boolean(B) term weighting is used, and Cosine(C) similarity measure is used for the Q-DIAC dataset, which is the Qur'an representation in Arabic with diacritics (case markings). The experimental configurations are summarized in Table 6.3.

Table 4.3: Precision and Recall Measures

		Cosine (C)			Euclidean (E)			Jaccard (J)			Pearson (P)		
		B	**TFIDF**	**F**	**B**	**TFIDF**	**F**	**B**	**TFIDF**	**F**	**B**	**TFIDF**	**F**
Q-NODIAC	**TP**	775	775	775	775	775	775	775	775	775	775	775	775
	FP	2	2	1	2	1	1	2	2	2	2	1	2
	FN	0	0	0	0	0	0	0	0	0	0	0	0
	P	0.997	0.997	0.999	0.997	0.999	0.999	0.997	0.997	0.997	0.997	0.999	0.997
	R	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
	F	0.999	0.999	0.999	0.999	0.999	0.999	0.999	0.999	0.999	0.999	0.999	0.999
Q-DIAC	**TP**	737	737	737	737	737	737	737	737	737	737	737	737
	FP	3	3	2	3	2	2	3	2	2	3	2	2
	FN	38	38	38	38	38	38	38	38	38	38	38	38
	P	0.996	0.996	0.997	0.996	0.997	0.997	0.996	0.997	0.997	0.996	0.997	0.997
	R	0.951	0.951	0.951	0.951	0.951	0.951	0.951	0.951	0.951	0.951	0.951	0.951
	F	0.973	0.973	0.974	0.973	0.974	0.974	0.973	0.974	0.974	0.973	0.974	0.974
Q-ENG	**TP**	661	661	660	661	660	660	661	661	660	661	660	661
	FP	12	12	10	12	10	10	12	12	10	12	10	12
	FN	114	114	115	114	115	115	114	114	115	114	115	114
	P	0.982	0.982	0.985	0.982	0.985	0.985	0.982	0.982	0.985	0.982	0.985	0.982
	R	0.853	0.853	0.852	0.853	0.852	0.852	0.853	0.853	0.852	0.853	0.852	0.853
	F	0.913	0.913	0.913	0.913	0.913	0.913	0.913	0.913	0.913	0.913	0.913	0.913
Q-ROOTS	**TP**	738	738	738	775	775	775	738	738	738	738	738	738
	FP	106	98	100	259	249	249	106	96	96	106	98	100
	FN	37	37	37	0	0	0	37	37	37	37	37	37
	P	0.874	0.883	0.881	0.750	0.757	0.757	0.874	0.885	0.885	0.874	0.883	0.881
	R	0.952	0.952	0.952	1.000	1.000	1.000	0.952	0.952	0.952	0.952	0.952	0.952
	F	0.912	**0.916**	0.915	0.857	**0.862**	**0.862**	0.912	**0.917**	**0.917**	0.912	**0.916**	0.915

4.3.3 Results and Analysis for Identical Verses

The results for the 48 experiments are shown in Table 6.3. We chose the F1 measure in order to provide a general measure of performance of the similarity configurations. We highlight the performance of each dataset below and then provide an overall comparison:

Q-NODIAC Representation

The Q-NODIAC representation provides the most accurate results when dealing with the Identical similarity class, with a perfect recall and almost perfect precision, and therefore the F1 measure being close to 1. The best result is obtained using C-F, E-TFIDF, E-F, and P-TFIDF combinations (indicated by 1 pair of verses classified as False Positive). All other combinations also perform almost as well. The difference is indicated by the number of False Positives (FP)s, in this case 2 pairs of verses are classified as FPs. The two FPs are the verse pairs [23:83, 27:68] and [22:62, 31:30][3]. Looking into the reasons for misclassification, we notice that the former pair includes a word that occurs twice in the verse 23:83 but occurs only once in 27:68. This is correctly captured using some of the combinations but missed with others such as the ones involving Boolean term weighting measure. The later verse pair, is a unique case, where the words are similar in both verses, however, the order is different. This is a unique case indeed, and perhaps, the only one of its kind in the Qur'an.

[3] 23:83 indicates Chapter 23, Verse 83 in the Qur'an

Q-DIAC Representation

The Q-DIAC performs considerably close to Q-NODIAC in terms of precision, however, the recall suffers due to the strikingly large number of verse pairs classified as False Negatives (FN)s. The verse pairs classified as FPs are the same as those in Q-NODIAC, apart from one additional verse pair i.e. [30:52, 27:81]. The reason for this verse pair being mis-classified in the difference in orthography of the words, which are not captured in the original dataset. By manual inspection we analyzed the verse pairs in the FN category, and we discovered that although the verses are actually similar, the orthography of the words, which is preserved in the diacritics as per the original manuscript causes them to appear different. For these experiments we preserved this in the original dataset. However, some of these orthographical cases may be removed without loss in any phonetic or linguistic characters of the word. Doing so, we expect that the Q-DIAC results would follow closely the results of Q-NODIAC dataset.

Q-ENG Representation

The Q-ENG representation, amongst all those approaches which use the raw Qur'anic representation, performs with a relatively low recall of around 85%. There is very little difference in either of the methods as indicated by the F1 measure which comes out to be the same for all. This representation reports the greatest number of False Negatives (FN) as indicated in the Table 6.3. This is a significant indicator especially when it comes to the identical verses. A FN implies that the verse pair which was expected to be classified as identical fails to do so. This has implications towards the number of translations available. It is clear that the identical verses in different occurrences within the Qur'an are being translated differently. To verify this we took a few sample cases and verified this implication. وَلَقَدْ يَسَّرْنَا الْقُرْآنَ لِلذِّكْرِ فَهَلْ مِن مُّدَّكِرٍ This verse is repeated in 54:17 and 54:22. The translation differs slightly (by one word) in the dataset we used for our experimentation.

The precision in the translation is a subjective issue. In the verses 54:17 and 54:22 the context may cause the first initial to mean as 'and' or 'but'. However, in another translation, the two verses are translated the same.

Analyzing the FPs, we discovered that in some cases, the English translation is the same but the Arabic terms used are different. This is significant when analyzing similar verses. Arabic is a rich morphological language. In particular the language of the Qur'an is considered to be precise and slightest variation or alteration in the arrangement or morphological manifestation of the word implies something significant, which often the translation fails to capture. E.g. in Table 4.4, the two verses are compared using the original text. The arabic text clearly distinguishes سحار and ساحر. The two words actually provide a different connotation, which is not captured by the English translation. These cases confirm this aspect of the translations that are available for the Qur'an. This can be used as a good measure for assessing the quality of translations. Another case is that of verse 37:80 [إِنَّا كَذَٰلِكَ نَجْزِي الْمُحْسِنِينَ] and verse 37:110 [كَذَٰلِكَ نَجْزِي الْمُحْسِنِينَ]. The word إِنَّا is well known in Arabic for a particle of extreme emphasis; this word is present in the former verse, however it is absent in the later. The english translation does not capture the difference in emphasis present in the arabic speech, as indicated by the highlighted word, in the context of these two verses.

Although, for this study we only used one translation of the Qur'an, available in English, our framework can prove useful in analyzing the quality of translations. An examination of those cases in other translations revealed that there are translations that distinguish clearly between near identical verses. This indeed is a reflection of the quality of the deliberation and effort that has gone into the translation process and thus may be measured to some extent.

Table 4.4: Analysis of false positive case (Q-ENG)

Verse	Verse Text(English)	Verse Text (Arabic)
7:112	And bring up to thee all (our) sorcerers well versed.	يأتوك بكل سحار عليم
26:37	And bring up to thee all (our) sorcerers well versed.	يأتوك بكل ساحر عليم

Q-ROOTS Representation

For the Q-ROOTS dataset, while recall is high, precision is quite low compared to the other datasets. In terms of the overall F1 measure, the Q-ROOTs and the Q-ENG datasets perform close to one another. However, the recall is much higher for the Q-ROOTs dataset compared to the Q-ENG dataset. In terms of the similarity measures, the pattern is not as consistent as with the other datasets. The C, J and P similarity measures perform closely. However, the Euclidean measure has a much lower F1 measure in comparison to the others. Interestingly, the recall for Euclidean comes out as 1. However, the precision is very low compared to the others. An investigation into this revealed some interesting results. An analysis of the FN cases showed that all 37 verse pairs are actually special verses which occur at the beginning of some chapters, and are special compositions of letters which are said to have no known roots or meanings. The Q-ROOTs dataset thus does not assign these terms any weight. The Euclidean measure, however, counts them as equivalent, even though the verse vectors are empty. Nevertheless, due to the raw nature of that similarity measure it computes them as equal. This explains the rather high number of FPs using the Euclidean measure, as compared to the other three. We manually inspected the verse pairs in this category to verify this. If we discount those verses, the measures would report equivalent. A future improvement to this measure could disregard those verses, or make a special case for the other similarity measures. If this is the case, all other methods would also report a perfect recall. However, for the initial experimentation, we preserved this result, because it indeed highlights the unique nature of those verses. Another interesting observation is that the Q-ROOTs returns the highest number of verse pairs with similarity value 1. This is indicative of high semantic similarity.

Overall Comparative Analysis

In general, we can conclude that using the Q-NODIAC representation provides the most accurate results when dealing with the high similarity class. In terms of the similarity methods used, the P-TFIDF configuration, i.e. the Pearson method, with TFIDF as the weighting measure and the C-TFIDF, the cosine method, with

TFIDF as the weighting method perform close and comparable. The analysis of the Similarity class where the verse pairs are identical is not sufficient to provide a reasonable estimate of the precision of the method that would perform the best overall. We therefore also look into some of the verse pairs that fall in the next similarity range to give us a reasonable estimate.

4.3.4 Results and Analysis for Verses with High Similarity

While the similarity method performance results are similar for the identical cases, this is not the case for the similarity ranges other than identical. This is shown in Figure 4.2. An investigation into the verse pairs retrieved with similarity values falling in the range of more than 0.9 and less than 1 reveals the numbers shown in Figure 4.2.

The Pearson method returns the most number of verses in that range. This is an important finding, as the results are meaningful. Verses with reasonably high similarity, which are expected to fall in this range are not falling in that range when using Cosine, Jaccard or Euclidean. We manually inspected some verses, which bear high semantic similarity. The Q-ROOTs dataset returned a large number of FPs, some of which are actually verses with a difference of only a single literal. When we looked into the similarity values of such pairs, and compared these values across the different experimental configurations, we observed that for a highly similar verse pair, the Pearson method returns the highest similarity values, whereas the similarity values returned by other methods are much lower. This pattern is reflective in Figure 4.2 which displays the number of verse pairs

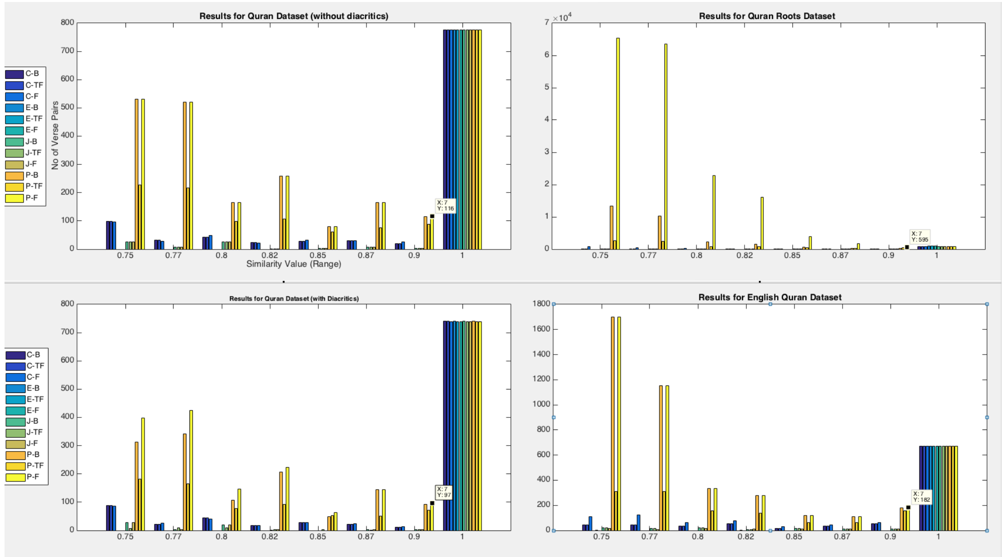


Figure 4.2: Results for all representations (Similarity Range vs. No of Verse Pairs)

in the similarity ranges from 0.75 to 1. As an example we took the verse pair [94:5, 94:6], which differs by a single literal. For both Q-DIAC and Q-NODIAC the values returned by Cosine, Jaccard and Euclidean are lower than 0.75, whereas for the Pearson method, the highest values are above 0.95. Therefore it may be concluded that Pearson performs best. As for the weighting measures the TFIDF measure with pearson returns the highest value. However, the performance of the term weighting measure may not be generalized from these instances alone and would require further investigation.

4.4 Discussion and Analysis

Our study, as outlined and described in this chapter, set out to quantify similarity between text segments, which is considered an important step in authorship attribution, corpus comparison, information retrieval, text mining and other fields. We attempted to provide an essential foundation by providing a means to choose an effective text similarity measuring scheme especially across cross-lingual texts of the same origin. We concluded that for the identical verses, the best results are obtained using the Q-NODIAC dataset, and all similarity measures perform comparably well. The Q-ROOTs dataset shows relatively low precision.

Challenges in measuring Text Similarity vs. Text Relatedness: It is worth mentioning that the notion of both text similarity and text relatedness are of strong importance when studied in the context of the Qur'anic scripture.The vector space model (VSM) or the bag of words approach imposes certain limitations and has several implications. The order of words is ignored which may be of value. The issue of synonymy is ignored. In addition, the issue of polysemy is not addressed because the VSM model only considers word forms. These limitations have been highlighted in previous research such as [70].

However, we feel that, the Qur'an is a text that deserves considerable attention when it comes to similarity at different levels. The results from this study will provide sound grounds to build upon and enhance the proposed similarity measures and methods with stem based methods to provide for more enhanced similarity classes. A future comparative study is planned to this end.

4.5 Conclusion

In this chapter, we investigated the application of various similarity measures to heterogenous representations of the Qur'an. The results focused on an analysis of the identical verses, delineating upon some interesting findings. As a natural next step, we aim to obtain some highly reliable similarity measures for non-identical verses, especially in the higher similarity ranges in order to establish ground truth for verse similarity in the Qur'an. We believe our study will prove to be a useful step in this direction. We aim to involve users and experts through established crowdsourcing approaches. We hope to devise an improved, hybrid similarity measure in the future for establishing a more precise estimate of semantic similarity. We also plan to extend our work to include other translations of the Qur'an and carry out a comparative analysis, as our study revealed some interesting findings that can help assess the quality of translations, and help standardize the existing translations.

Chapter 5

Crowdsourcing Framework for Linked Data Management

إِن يَنصُرْكُمُ ٱللَّهُ فَلَا غَالِبَ لَكُمْ ۖ وَإِن يَخْذُلْكُمْ فَمَن ذَا
ٱلَّذِى يَنصُرُكُم مِّنۢ بَعْدِهِۦ ۗ وَعَلَى ٱللَّهِ فَلْيَتَوَكَّلِ ٱلْمُؤْمِنُونَ

If Allah should aid you, no one can overcome you; but if He should forsake you, who is there that can aid you after Him? And upon Allah let the believers rely.

— *[Al-Quran, Ale-Imran,3:160]*

Chapter Overview[1]

This chapter presents the design and development of a semantically enriched task management and workflow generation mechanism applied to the domain of ontology verification, more specifically in the domain of linked data. Verification of ontologies particularly in the domain of linked data is a challenging and complicated task, which requires human expert knowledge on one hand and is time consuming for the humans on the other hand. The rationale for our proposition implies that employed semantics and automated workflows will yield better workflow management for crowd based tasks and can thereby help solve more complex computational problems, especially in big data sets such as the LOD. We devise a task management, workflow specification and automation mechanism that aims to make this process more efficient. This framework forms the foundation of the crowdsourcing module which is utilized in the case studies presented in chapter 6.

5.1 Introduction

This research focuses on crowdsourcing, which employs a large number of human beings rather than or in addition to software or algorithms as information processing units, since humans outperform software in certain information centric tasks ranging from labeling or segmenting images to handling more complex tasks such as finding the same entities in different data sets. Crowdsourcing is task-oriented

[1]This chapter is published as:
Basharat, A., I. Budak Arpinar, Dastgheib, S., Kursuncu,U., Kochut,K., and Dogdu,E. "Semantically Enriched Task and Workflow Automation in Crowdsourcing for Linked Data Management", *International Journal of Semantic Computing*, Vol. 8, No. 4 (2014) 1-25.

and hence specification and management of tasks play a critical role. We propose design and development of a semantically enriched task management and workflow generation mechanism applied to the domain of ontology verification, more specifically in the domain of Linked Open Data (LOD) [36, 47]. Verification of ontologies particularly in the domain of LOD is a challenging and complicated task, which requires human expert knowledge on one hand and is time consuming for the humans on the other hand. The rationale for our proposition implies that employed semantics and automated workflows will yield better worflow management for crowd based tasks and can thereby help solve more complex computational problems, especially in big data sets such as the LOD. We devise a task management, workflow specification and automation mechanism that aims to make this process more efficient. The work presented in this chapter is an extension of our recent work [23].

5.1.1 Research Context

Recently, crowdsourcing has emerged as a major data collection and problem-solving paradigm on the web. Many well-known applications and tools including the Wikipedia, YouTube, Flickr etc. employ various crowdsourcing techniques. Database community also shows increasing interest in this growing field. Berkeley's CrowdDB [68], Stanford's sCOOP [151], and MIT's Qurk [129] are some examples of crowdsourcing-based query processing systems. Many crowdsourcing applications are built using private or public platforms. Amazon Mechanical Turk (AMT) [1] and CrowdFlower [11] are most prominent of these platforms. In our review of crowdsourcing literature, we have observed the need for the provision for defining better means of automatically generating task based workflows, and their data and control-flow interdependencies. In addition, there is currently very limited support for well enriched and semantically enhanced definitions of tasks and worker profiles. We use this as our motivation by contextualizing the need for semantically enriched task and workflow automation to the domain of LOD management.

In a parallel stream to crowdsourcing, Linked Data garners significant attention from academia and industry [36, 47]. There has been a recent and tangible growth in RDF and OWL triples published on the (semantic) web in accordance with the Linked Data principles and best practices [93, 104]. In 2007, the W3C LOD project began publishing legacy web corpora under Linked Data principles. This resulted in rich datasets; most prominently the DBpedia corpus extracted from semi-structured Wikipedia articles. Thereafter, Linked Data adoption spread to various corporate entities, including BBC, Thompson Reuters, and the New York Times joining the effort and exposing information as Linked Data. More recently, various governmental agencies have begun disseminating various public corpora as Linked Data, beginning with commitments from the US and UK governments, and spreading to various other governmental bodies.

As of September 2011, LOD included 295 data sets consisting of over 31 billion RDF triples, which are interlinked by around 504 million RDF links. Despite LOD's sheer size, [93] and [94] report varying levels of quality in LOD data sets. In our opinion, the most important quality issue is the accuracy of links within and across the LOD data sets. If some of these links are not accurate or missing, this will defy the main objective of the LOD. For example, *owl:sameAs* relation is used to relate two co-referent resources that talk about the same real-world entity. In [79], the authors take an initial set of 58 million *owl:sameAs* triples extracted from 1,202 LOD domains. They then employ four human judges to manually inspect 500 links sampled from the full corpus. Their experiments found that only approximately 51% of *owl:sameAs* relations were deemed correct. [93] reports a more optimistic result in terms of accuracy of *owl:sameAs* links in LOD.

5.1.2 Motivating Examples

Our motivation for this work is two-fold: (i) enabling a more systematic task and workflow specification and management using (semantic) workflows, and (ii) providing more efficient means of Linked Data management tasks for the LOD using crowdsourcing. Task management and specification is a critical component currently lacking in the current crowdsourcing platforms such as Amazon Mechanical Turk (AMT). For instance, using AMT to devise a task for the verification of links in the LOD is neither efficient nor scalable. Hence, we are motivated to investigate how crowdsourcing may be semantically enriched to make the process of task management, in particular for those tasks pertinent to the data management on the LOD, more efficient.

Linked Data acquisition and verification is a challenging task, which requires human knowledge. The massive growth of the LOD has made this process even harder. For example, verifying the screenwriters of movies in a movie ontology from the LOD involves human knowledge in domain of movies, and the human capability to distinguish and disambiguate the entities. We believe that such verification tasks can greatly benefit from the idea of crowdsourcing. Ideally a machine-based algorithm would identify any cases of entity disambiguation, however, only a human could verify the correctness of some disambiguations based on available data and also his background knowledge. It is an extremely difficult task to check every triple out of billion of triples in the LOD cloud. High value links within a specific domain which is of particular interest need special attention, e.g., mutations causing certain diseases in the biomedical domain would need verification by domain experts. We outline a few more motivating examples below as potential basis of our investigation.

Geographical Locations

As an initial example for motivating our research, consider the following geographical relation in the LOD: the city of Paris is referenced in a number of different LOD data sets, ranging from OpenCyc to the New York Times. [79] finds that (1) *dbpedia:Paris* is asserted to be sameAs both (2)*cyc:CityOfParisFrance* and (3)*cyc:ParisDepartmentFrance* (and five other URIs). Yet, OpenCyc explicitly states that (1) and (3) are distinct. In our opinion, crowdsourcing empowered with human-machine workflows can facilitate verification of accuracy of such geographic links (and beyond), in addition to incorporation of new *owl:sameAs* links. In this way, LOD can achieve its full potential as the core of the semantic web. LOD also has a potential to accelerate widespread adoption and creation of semantic web, and also to become a meta-index for the traditional web (e.g., web documents).

Publications

As another example, consider two large-scale data sets on Computer Science publications, namely DBLP and CiteSeer, which are in the LOD cloud. A true potential of these two data sets can only be achieved if they are properly linked through common authors, publications, etc. A fully automated approach (such as Silk [198]) can cross-link the same entities with reasonable accuracy unless the reconciliation decisions are very complex. Consider an author named "F.S. Arpinar" in DBLP and "Sena Nural" in CiteSeer. Do they represent the same Computer Scientist? Their names look very different syntactically, their homepages or e-mails are not listed in both data sets, and also their papers and co-authors do not have any commonalities. In this situation, a crowdworker can do better than a fully automated approach by analyzing surrounding entities in LOD and other web sources to capture affiliations, co-authors, etc. of these two persons. He can also use his background knowledge (if relevant) to recognize that both names are Turkish, and Sena is a female name. This may provide an insight that the last names might correspond to the ones before or after a marriage. In fact, these two names belong to the same person, and in our opinion, only a human can offer a conclusive reconciliation decision in this case.

Movies

If you search for the movie "Ice Age" on the web, you will not end up finding a single one. A German movie "Ice Age" was made in Germany in 1975 by "Peter Zadek", while "Ice Age" computer animated series have been produced since 2002 by different directors. For example, "Chris Wedge" directed, and "Carlos Saldanha" co-directed the first one in 2002, and later "Saldanha" made the "Ice Age: The Meltdown" in 2006. We want to verify if every "Ice Age" movie is associated with the correct director in Linked Movie Data Base (LinkedMDB) [22], an open semantic web database for movies.

5.1.3 Main contributions

We believe that our work is the first that promotes the use of automated tasks and workflows in crowdsourcing for the purpose of linked data management. We propose a generic and reusable mechanism for crowdsourced task specification, generation, publishing, reviews and updates for the purpose of linked data creation and verification [Section 5.2]. The components of our task and workflow management are generic enough to be adopted for other task classifications with minor changes to the workflow. Further, we have introduced the notion of ontology based task specifications and worker profiles for the purpose of improved task selection and workflow generation [Section 5.2.4]. Workflow management also benefits from our approach since traditional workflow systems usually omit web scale crowdsourcing tasks and their inclusion as computational units (i.e., services) in complex workflows. In addition, our approach facilitates improving link quality (i.e., triples) in the LOD by employing crowdworkers for link addition (e.g., addition of *owl:sameAs* links) and verification (e.g., checking accuracy and correctness of triples surrounding a particular resource) [Section 5.3]. We also provide insights into how our approach compares and contrasts with existing systems and approaches [Section 5.4]. Finally, we share some envisioned extensions to our current work particularly in context of how crowdsourcing may benefit the domains of query processing and text mining [Section 5.5].

5.2 Crowd-based Task and Workflow Management Architecture for the LOD

Fig. 5.1 shows the high-level architecture of our crowdsourcing-based task specification and workflow management engine. It facilitates building LOD management workflows by the ontology experts and knowledge engineers.

5.2.1 Architecture Overview

The complete execution of a task lifecycle using any particular crowdsourcing system involves several general stages: (i) a task is defined, with input parameters and the needed data; (ii) the task is published over the crowdsourcing platform; (iii) the tasks is searched for or discovered by the crowdworkers (via search or direct notification); (iv) the crowdworkers perform and submit the task(s); (v) the submitted tasks then need to go through a review process for obtaining the submissions; (vi) the submissions are either accepted or rejected; (viii) the accepted

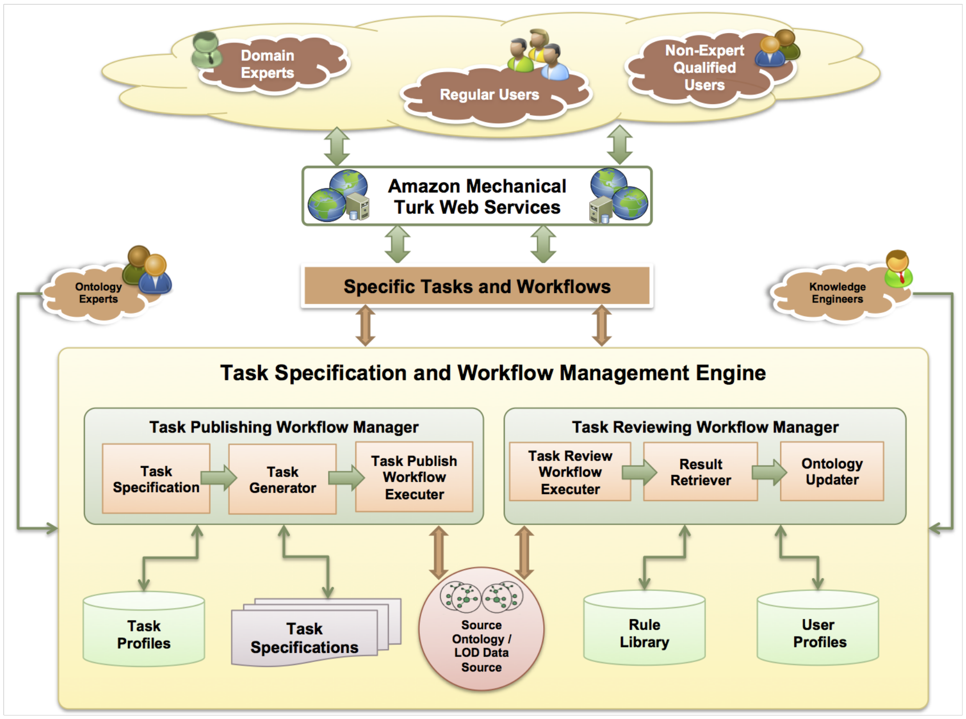


Figure 5.1: High-level Architecture of the Semantics-Enriched Workflow and Task Management Engine

submissions are further processed for retrieving and synthesizing relevant results.

We have devised the key component of our architecture as the task specification and management workflow engine, which in turn consists of several components detailed in subsections below. However, the most signification of these are (1) *Task Publishing workflow manager*, and (2) *Task Reviewing workflow manager*. These two sub-components correspond to the life-cycle stages of the tasks performed via a crowdsourcing system. Fig. 5.1 also shows the different stages and components involved in task management and workflow engine in detail.

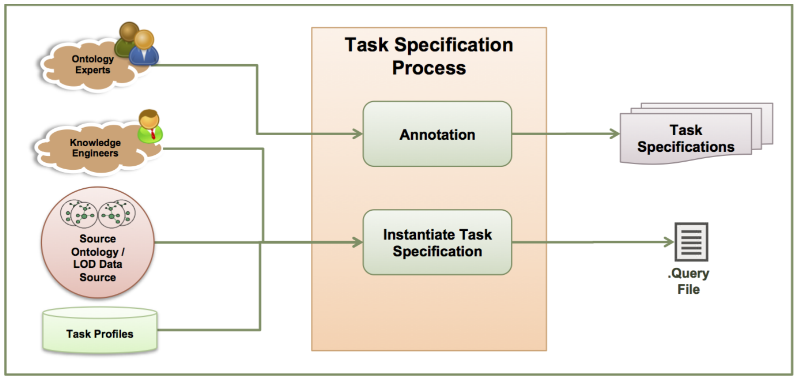


Figure 5.2: The Process of Task Specification

5.2.2 Task Generation and Publishing Workflow

The key objective that we achieved through this engine has been the ability to generate dynamic tasks and worker workflows for any provided ontology or an LOD data source based on some rules specified by the ontology expert.

The process of task specification is shown in Fig. 5.2. The specification process results in the creation of two important elements that play a key role in the execution of the task life cycle: (1) *A Task Specifcation*, and (2) *A query file*. The task specification is the definition which includes all the essential parameters needed for successful completion. More details are included in Section 5.2.4. A sample task specification is shown in Table reftable:t-spec1. The tasks are classified according to various task parameters such as *Task Eligibility*, *User Type*, *TaskType*, *Task ID*, *Date of Creation*, *Date of Modification*, *Created By*, *Modified By*, *System Verification Status*, *Review Status*, *Review Approval Status* and *Reviewed By*. Note that only most essential parameters have been specified in Table reftable:t-spec1.

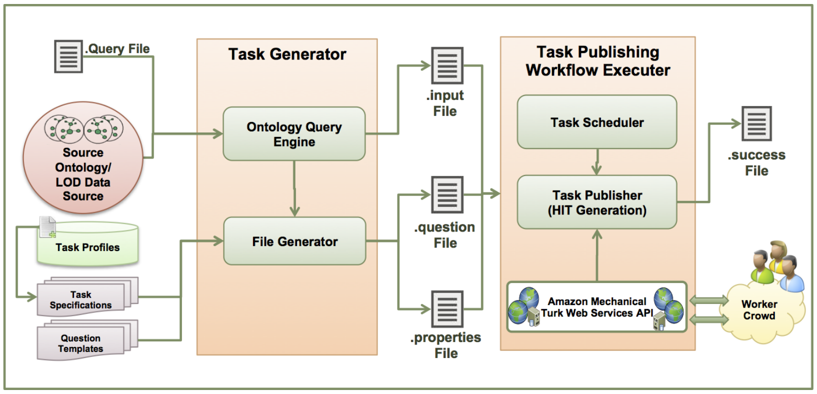


Figure 5.3: The Process of Task Generation and Publishing

We created a task generator, which dynamically generates tasks based on the provided task specifications, resulting in the creation of all the necessary elements (primarily including the .input, .question and .properties files), which are then used to publish AMT compatible HITs (Human Intelligent Tasks) dynamically and seamlessly on the AMT platform using the task publishing workflow executer. The process of task generation is detailed upon in the Fig. 5.3. We provide task specifications to the task generator which have been created based on the pre-existing task profiles, which have been annotated by a human expert. Along with the task specifications, certain question templates are also provided. The question templates correspond to those supported by the AMT platform. However, we customize those templates to include parameters for the support of automated generation of the question files needed for publishing HITs on the AMT.

One of the key elements specified in the task specification is the data source. There is an optional query component to the task specification in which a SPARQL

Table 5.1: A Sample Task Specification (Review Type)

Task Profile Parameter	Parameter Value
TaskID	T0001
TaskName	ReviewMovieDirector
TaskType	Review
TaskDescription	Please Review the if the Movie Director name mentioned below is correct for the given movie
TaskQueryTemplate	./queries/query2.txt
TaskQueryLimit	10
TaskUpdateTemplate	Null
TaskDataSourceType	endpoint
TaskDataSource	http://data.linkedmdb.org/sparql

query may be attached to specify the precise elements of the LOD to be verified. The data to be fed as input to the task may be queried directly from a live LOD source using a SPARQL endpoint. We elaborate on this using the movie example mentioned in Section 5.1.2. In our Movie example, we have used Linked Movie Data Base (LinkedMDB) which is a movie data source on LOD with a SPARQL endpoint (http://www.linkedmdb.org/snorql/). The query shown in Listing 5.1 is formulated to verify if the movie director name is correct for a given movie released on a given date. A reference to this query file is specified in the task specification. When this is provided to the Task Generator, along with the query file, the data file needed for processing HITs on the AMT is automatically generated using the Ontology query engine and the file generator included in the Task Generator module. Each task to be published over AMT has an associated data file (.input file). The resulting data file contains the data obtained by querying the ontology or the live LOD source specified in the task specification. Table reftable:t-data shows the data file generated by using the query shown in Listing 5.1. The number of the data variables (no of rows) in the data file corresponds to the limit specified in the query file. It may therefore be controlled by altering the parameter. The number of the HITs generated on the AMT for each task group depends on this number. For the purpose of our experimentation, we kept this number to be 10 for all cases.

Along with the data files, the task generator automatically creates other files essential for task execution over AMT. These include the property files for task parameter specification over AMT (e.g., task title, amount of the reward to complete the task, assignment duration, number of assignments, etc.), and the question file in XML format which includes the final question (task) template to be supplied to the crowdworkers via AMT. As seen in Fig. 5.6, we not only ask for the crowdworkers' verifications, but also for their confidence level. A snippet of the parameters in the property file corresponding to the task specification (in Table reftable:t-spec1) are shown in the Table reftable:t-spec2. Note the Serial serves as a key that helps identify different HITs belonging to each Task Group (A task group basically refers to many HITs of the same Task). Note that the values presented here are for the sake of illustration. The actual encoding format for both the tasks specifications and the property file is different.

The task scheduler manages publishing and review of tasks and takes care of the task dependencies. The task publisher uses the AMT API and the required data,

Table 5.2: Sample Task Data Input Generated from the Query Processor

Serial(HITid)	MovieTitle	MovieDirector
1	"My Wrongs 8245 - 8249 and 117"	"Chris Morris"
2	"Four Lions"	"Chris Morris"
3	"The Greatest Show on Earth"	"Cecil B. DeMille"
4	"The Cheat"	"Cecil B. DeMille"
5	"The Ten Commandments"	"Cecil B. DeMille"
6	"Samson and Delilah"	"Cecil B. DeMille"
7	"Reap the Wild Wind"	"Cecil B. DeMille"
8	"King of Kings"	"Cecil B. DeMille"
9	"The Squaw Man"	"Cecil B. DeMille"
10	"Male and Female"	"Cecil B. DeMille"

question and property files generated by the task generator to publish the tasks as HITs on the AMT. A success file is generated if this process is successful. This file includes all the HIT identifiers, which are used to supply to the crowdworkers. The HITs are also discoverable over the AMT search. However, for the purpose of evaluation, we supplied these accessible HIT links to the prospective crowdworkers.

Listing 5.1: Query Attached with Task Specification

```
SELECT ?MovieTitle ?MovieDirector ?MovieDate
WHERE {
        ?movie_uri rdf:type movie:film;
                dc:title ?MovieTitle;
                dc:date ?MovieDate;
                movie:director ?director.
        ?director movie:director_name ?MovieDirector.
```

Table 5.3: Sample Task Assignment Properties

Assignment Property	Property Value
Annotation (HITid)	${Serial}
Title	ReviewMovieDirector
Keywords	Review, Movies, Director
Reward	0.00
HitLifeTime	25920000 *[The time duration for which the HIT would remain active on AMT]*
Description	Please Review the if the Movie Director name mentioned below is correct for the given movie
Assignments	20
AssignmentDuration	240000 *[The time allowed to perform a HIT]*

```
    }
LIMIT 10
```

5.2.3 Task Reviewing and Results Processing

Once the tasks have been submitted by the crowdworkers, the results are retrieved, processed and the retrieved results are updated into the ontology (if needed). Fig. 5.4 shows the process of task reviewing and retrieval. The task scheduler once again schedules the review workflow. The review processor module is responsible for the analysis of HIT result submissions. The key process performed during this stage is deciding whether to accept or reject any submission by the crowdworkers. The submissions may either be accepted by default or rejected based on missing or vague submissions. Using the HIT success files from the task generation workflow, HIT results are retrieved and analyzed. For the initial evaluation, all HIT submissions from crowdworkers are accepted by default and no worker qualifications are enforced, since the experimentation is intended for closed worker environment over AMT Sandbox. Once the results are retrieved, they are processed using the Result Processor. Since the same tasks are assigned to multiple workers, multiple submissions are possible. A confidence-weighted measure based on majority of votes is used to decide the final outcome for the task. Those tasks, where the minimum confidence threshold is not met are withheld for further review or another round of evaluation. For the approved tasks, the results are provided to the ontology population engine, which uses the SPARQL update protocol to update the ontology as shown in Fig. 5.5. For consistency and verification, a separate snapshot of the ontology is maintained. As part of the update process, certain changes may also be reflected in the task profiles or task tpecifications. This includes updating certain status parameters such as task completion status, so any other task in task sequences may be triggered or incompleted tasks may be scheduled. Status also needs to be maintained for those elements of the LOD which have been reviewed.

5.2.4 Ontology based Task Profiles

One of the key aspects of the CrowdLink platform is the provision of ontological repositories for maintaining user and task profiles, their classifications and assignments. Task and user classification is carried out using rules and reasoning based on the knowledge defined in ontologies for better workflow modeling and task recommendation, coupled with notifications and recommendations to workers upon

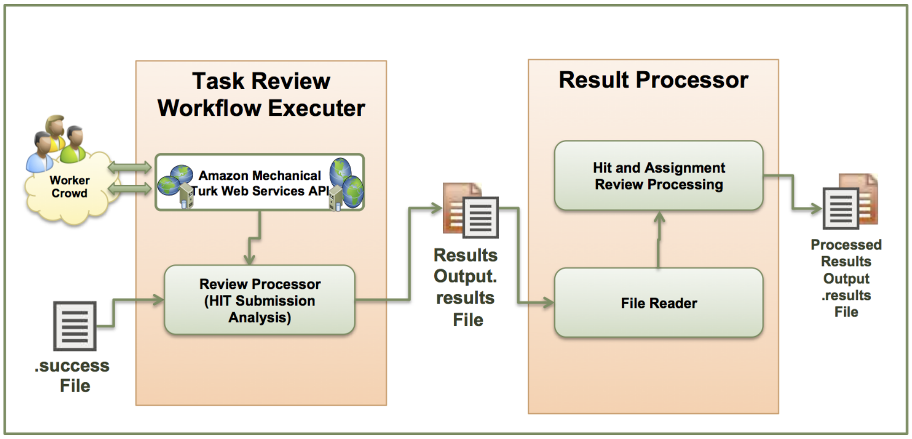


Figure 5.4: The Process of Task Review and Results Retrieval

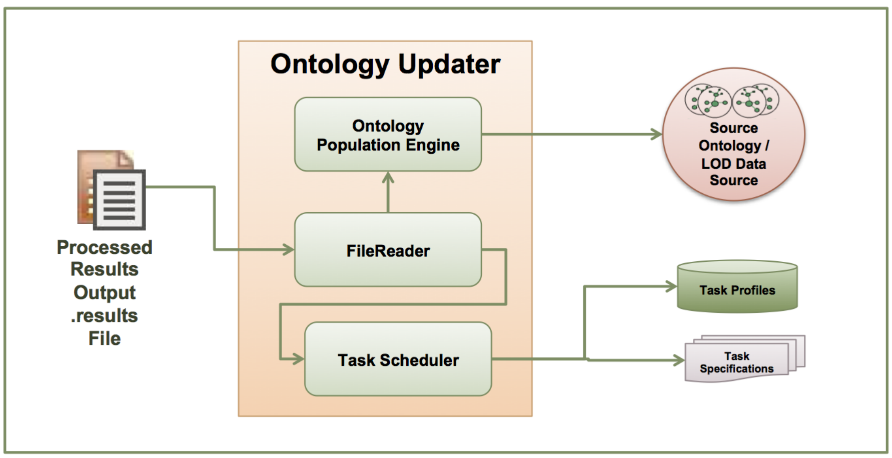


Figure 5.5: The Process of Updating the Ontology from the Processed Results

relevant task and workflow availability.

In order to create dynamic workflows, we have created an ontology driven mechanism for task specification. Tasks are broadly classified as verification and creation tasks, where the former refers to those tasks that require the crowd-workers to verify the correctness of the existing data in the LOD. For example, crowdworkers need to decide if the director of the "Ice Age" movie which was made in 1975 is "Carlos Saldanha" (which is not correct). The creation tasks refer to those tasks, which require the crowdworkers to enter new knowledge. For example, the crowdworkers need to enter the correct director name (e.g. "Peter Zadek") or add the location in which a given movie was made (e.g., "Germany"). Tasks specifications are created at different levels of granularity, and more task specifications can be made available. A Task profile repository (shown in Fig. 5.1) includes generic schema definitions (based on OWL/RDF based Ontologies) from which task specifications are instantiated from. Task specifications are instances that become input to the task generation engine. The tasks are classified according to some key task categories in the domain of LOD management including, but not limited to:

- Knowledge acquisition (acquiring missing facts and information),
- Knowledge verification (for correctness, consistency, and relevance – verify provenance information or any other verification parameter),
- Ontology evolution (accommodating structural changes, and schema level changes).
- Entity disambiguation (identifying ambiguous links).

Our system provides some templates for a given task description or information request. Tasks are not limited to simple identification assignments, but other complex assignments such as entity disambiguation. Currently, we provide specifications for both instance level and schema level tasks. Further details of some tasks used in the experiments for the purpose of evaluation are presented in the Evaluation Section (Section 5.3.1).

Task specifications can either be generated in an automated manner based on some predefined task templates or new task specifications can be created by ontology experts. In addition, a task specification can be annotated with relevant user profiles that may include task specific requirements such as worker qualification, expertise or experience required to perform the task.

5.2.5 Cooperative Workflows

We have also provided the provision for cooperative workflows that require different tasks to be performed in succession by the machines and the humans. Some tasks such as entity disambiguation represent this class of tasks.

For acquiring new *owl:sameAs* links between entities, tools such as Silk [198] can be used to identify candidate entities for potential *owl:sameAs* links. We can define a threshold for identifying those links, discovered by external link discovery tools such as Silk, that may need verification from the crowdworkers. We then present these to the crowd and verify the results. Another example of the coorperative workflow could be whereby there may be some missing data in the LOD, and the crowd may be engaged to perform new knowledge creation task to fill in for this missing data. Once this goes throughout the task management lifecycle, and is updated into the source, it becomes candidate for verification. Such a workflow may be iteratively defined using the task profiles available, whereby, respective task specifications would be instantiated from these profiles at the appropriate stages.

The definition of the such cooperative workflows require either task sequences or dependency rules to be defined. Tasks may be published in parallel if they don't have dependencies on other tasks. For those which require the completion of other tasks will have a dependency rule stated in the task specification. The role of the task scheduler is to resolve these dependencies and publish relevant tasks at relevant times. The task expiry will determine when to evaluate the status of completed tasks. Any such rules may be defined in the rule library, which is maintained separately from the task profiles.

5.3 Experimental Evaluation

In order to assess the feasibility of our approach, we have conducted an experimental evaluation using the AMT. Note that this experimentation is limited to triple creation and verification tasks in the LOD.

5.3.1 Details of Experimentation

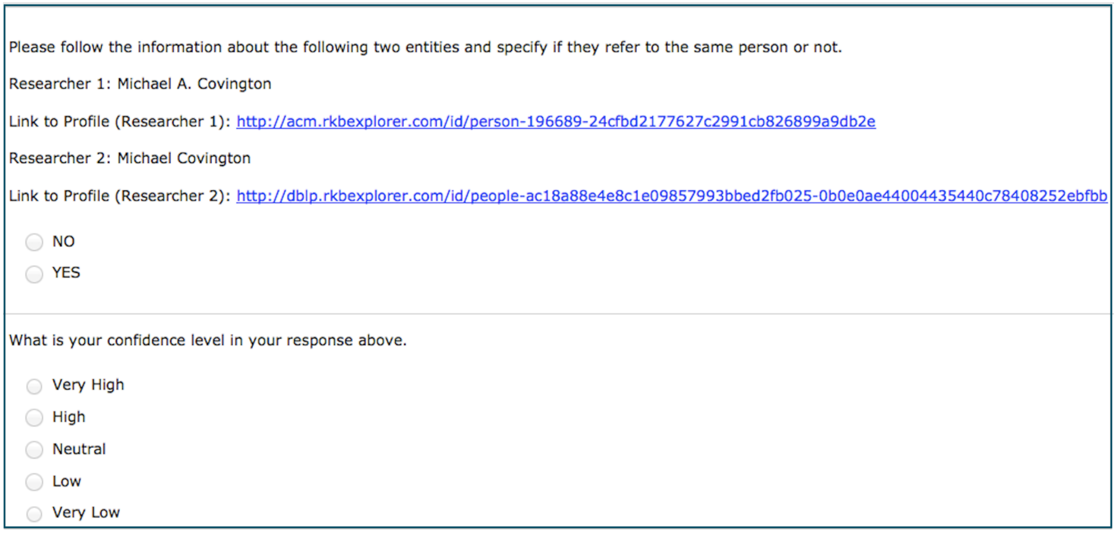
Task Classifications used for Evaluation

In our experiments, we classify HITs into two broad categories, and then into finer subcategories as follows: All HITs belong to either *verification* or *creation* categories. Verification tasks use crowdsourcing for verification of existing triples in the LOD. The creation tasks facilitate introducing new knowledge (i.e. new triples) into the LOD. Verification tasks can be further classified as follows: (1)

property verification, (2) entity verification, (3) schema-level verification, and (4) *owl:sameAs* verification.

In these tasks, workers check the validity of an existing property connecting two entities, or they check if one or both entities are valid for a given property in the LOD. In addition to the entity-level verification, concepts and properties at the schema-level can be scrutinized by human computation for their validity as well. Also, the use of *owl:sameAs* enriches the LOD by declaratively supporting distributed semantic data integration at the instance level. Primarily, it links the same real-world entities with different identifiers within a single or across different data sets. As mentioned earlier, empirical studies conducted on *owl:SameAs* reported various irregularities of its usage in the LOD [55, 79, 93]. Therefore, we consider verification of *owl:sameAs* relations an important task in our experimental study.

Creation tasks in the LOD can be classified similarly: (1) property creation, (2) entity creation, (3) schema-level creation, and (4) *owl:sameAs* creation. In this category, crowdworkers can introduce new properties, entities as well as schema-level elements in their HITs. Again, we treat *owl:sameAs* as a separate category because of its importance in cross-linking various data sets in the LOD. An example of the creation tasks involves disambiguating entities across multiple data sets in LOD, and establishing new *owl:sameAs* relations between these entities. Fig. 5.6 illustrates disambiguation of two researchers from ACM and DBLP data sets respetively. Note that, since we are not authorized to introduce new triples in the existing data sets in the LOD, the new triples are stored externally, as a tempo-


Please follow the information about the following two entities and specify if they refer to the same person or not.

Researcher 1: Michael A. Covington

Link to Profile (Researcher 1): http://acm.rkbexplorer.com/id/person-196689-24cfbd2177627c2991cb826899a9db2e

Researcher 2: Michael Covington

Link to Profile (Researcher 2): http://dblp.rkbexplorer.com/id/people-ac18a88e4e8c1e09857993bbed2fb025-0b0e0ae44004435440c78408252ebfbb

NO
YES

What is your confidence level in your response above.

Very High
High
Neutral
Low
Very Low

Figure 5.6: A Human Intelligent Task (HIT) in AMT

rary solution. In addition, we define tasks in both categories as *open* tasks, which means that the crowdworkers are allowed to use their own background knowledge as well as (reliable) web-based resources for triple verification and creation. Note that the workers also indicate their confidence level in completing assigned tasks as illustrated in Fig. 5.6. These confidence scores are used for consolidation of worker responses at a later stage. If conflicts (e.g., a tie in a verification task) cannot be resolved automatically, they can be forwarded to a superior level of workers who have better expertise and experience in the creation or verification tasks in a particular domain.

Results

In our evaluation, we had 100 sample queries (for Groups 1 – 10, as shown in Table reftable:task-gps), which involve mainly knowledge acquisition and verification tasks for datasets from US Census data, and Linked Movie Database. Table reftable:task-gps also lists various task groups, where each group contains ten individual tasks. We have recruited ten graduate students across UGA campus for completing these tasks (except task group #11 on entity disambiguation). It is worth mentioning here that crowdsourcing generally involves a large group of possibly unknown and even untrustworthy humans over the web, often using incentives such as micropayments or "fun". However, for the purpose of early experimentation and evaluation we restircted our user group to selected number only. Ideally, it should scale to large number of crowdworkers which we aim to do in future evaluations. Note that we have used AMT Sandbox, which doesnot require any payments for this evaluation. The number of responses collected during this evaluation is also given in Table 5.4.

5.3.2 Measure of Quality

In order to evaluate the effectiveness of our methods, we use the traditional metrics of precision and recall. More specifically, we utilize the definition of these metrics and adapt them for triple creation and verification tasks. We compare, for each triple, crowd provided subject, object or predicate against the Gold Standard (GS) (or the ground truth), which provides matching/non-matching information for each triple. Specifically, we compute (P)recision and (R)ecall, which are defined

Table 5.4: Eleven Task Groups Used in the Experimental Evaluation (G: Group ; P: Participants ; R: Responses)

G#	HIT Group	#P	#Rs	#Responses/Part.s
1	ReviewRiverNames	10	19	1.900
2	AcquireRiverNames	9	69	7.667
3	ReviewRiverLength	7	61	8.714
4	AcquireRiverLength	7	51	7.286
5	ReviewCounties	6	60	10.000
6	GetCountySchools	5	33	6.600
7	ReviewMovieInfoLink	4	33	8.250
8	ReviewMovieDirector	7	68	9.714
9	ReviewHeadquarterName	5	47	9.400
10	GetStateNameforRiver	6	53	8.833
11	EntityDisambiguation	20	127	6.350

as follows: We consider as *true positives (tp)* all cases where both the GS and the crowdworkers select a triple element (i.e., subject, object or predicate) (*correct result*). *False positives (fp)* are the cases where the crowdworkers select an element, which is not considered correct by the GS (*unexpected result*), and *false negatives (fn)* the cases where the crowdworkers do not select an element that is correct in the GS (*missing result*). Then, Precision is defined as $P = tp/(tp{+}fp)$, and Recall as $R = tp/(tp{+}fn)$. Fig. 5.7 illustrates the precision and recall for eleven task groups illustrated in Table 5.4. Note that each group contains ten tasks (total 100 tasks except *EntityDisambiguation* group), and precision and recall are averaged for these tasks in each group. The precision is generally good, and above 80% in four task groups. The recall is also above 80% in all task groups. However, the precision is low in certain task groups such as #2 (32.5%), which asks for

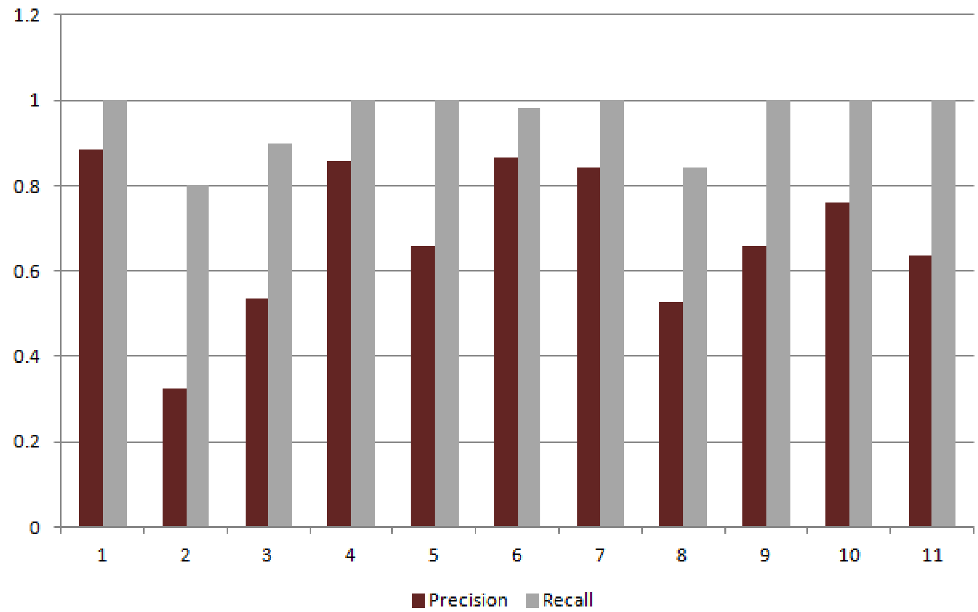


Figure 5.7: Evaluation Results for Eleven Task Groups

number of public high schools in a particular county. The number of schools can be listed slightly differently in county web page and elsewhere, and most likely crowdworkers had difficulty in judging about the most up to date information.

Note that no other existing systems are exactly comparable to our proposed system, as CrowdLink encompasses the following two primary aspects of linked data management: (i) linked data creation, and (ii) linked data verification. CROWDMAP [163], a system that focuses only on entity disambiguation task of linked data verification was published recently. We plan to conduct a comparative study of CrowdLink and CROWDMAP on entity disambiguation aspect in the near future.

5.3.3 Analysis

After execution of the experiments, collected result files contain various information about tasks such as HITid, title, deadline of the task, reviewStatus, workerid, answers, and confidence scores. HITid allows the system to provide a task to a worker only once. Since the AMT is an open platform to the public, people can search for a job that is suitable to their qualifications. For this reason, each answer needs to be reviewed and approved to be considered as a valid answer to make sure whether it is supplied by a qualified person with a confidence attribute. Workerid is another important attribute, because data processing filters answers based on workerid for analysis purposes. A portion of the result files generated automatically by our system is presented in Table 5.5, which contains the attributes mentioned above and the Gold Standard for each task, which is added at the analysis stage. 10 sample answers for different tasks have been presented in the table, and HITid and Workerid uniquely identify each response with their confidence levels in a result file.

In Table 5.5, we observe that the Gold Standard has more than one correct answer such as "TX, CA" for the first task, while most of the others have only one. Therefore, there is always a possibility that workers will not be able to give all correct answers in the Gold Standard. Also, there is also a likelihood that there might be tasks that no one will be able give the correct answer for, particularly the tasks with only one correct answer, and those situations give us false negatives. The effect of false negatives on the analysis is on recall value, and having false negatives reduces this value. In our experiments, recall values of most of the tasks

Table 5.5: Attributes and Answers in a Result file for GetStateNameforRiver Task

HITid	Workerid	Confidence	Answer	Gold Standard(GS)
id1	id1	High	Texas	TX, CA
id2	id2	Very High	Ohio	California
id3	id3	Very High	California	California
id4	id4	High	California	California
id5	id5	Very High	California	California
id6	id6	High	Pennsylvania	PA, MD, CA
id7	id7	Neutral	California	California
id8	id8	High	Oregon	Oregon
id9	id9	Very High	California	California
id10	id10	Very High	California	California

are 1, and others are above 80% and this shows that our approach to accomplish LOD tasks using crowdworkers has provided promising results.

Based upon the answers collected for the set of 10 tasks and the entity disambiguation task, we see a consensus among workers on the answer(s) in Gold Standard. From Fig. 5.7 above, we can see the precision values for each task and only two of them are below 60%, and four of them are above 80%, thus this indicates that workers have a common ground on their answers. For the tasks with lower precision, we believe that these tasks have been accomplished by workers who did not have confidence in their answers. Therefore, including those with lower precision, the precision values of each task will improve even better when we account confidence level value of the worker in the computation.

5.3.4 Confidence based Analysis for Entity Disambiguation Tasks

In the 11th task group, we published several entity disambiguation tasks from the LOD domain. Our objective was to reconcile same real-world entities, which can be represented through different identifiers (i.e., URIs) not only within a single data set, but also across multiple data sets. For experimentation, we used ACM and DBLP data sets, which contain large-scale data about researchers, their publications and their co-authors. We restricted ourselves to the domain of Computer Sciences.

A sample set of candidate entities extracted from ACM and DBLP datasets are given in Table 5.6. Due to space constraints, the corresponding entity, publication and co-author URIs cannot be depicted in the table. In this experimentation, 20 Computer Science undergraduate students were asked to identify if the given pairs of candidate entities represented the same researcher. Their confidence scores were also solicited as part of their responses using the AMT. AMT HITs also contained the URIs of the surrounding entities, such as co-authors and publications in the LOD. Also, the tasks were designed in a manner that multiple crowdworkers could submit assignments for the same task. We utilized this multitude of responses, weighted by their confidence scores to determine an overall measure of accuracy for the disambiguation tasks.

As indicated in Fig. 5.7, the overall precision and recall are comparable to other task groups in our experimentation. In general, the more difficult queries would be expected to have lower precision. An interesting result is obtained when we grouped the results by TaskIDs and WorkerIDs along with not only the obtained precision and recall values, but also the average confidence measure reported by the workers as shown in Fig. 5.8 and Fig. 5.9. The confidence measure was quantified by mapping the range (0-1) to levels Very Low(0) to Very High(1). The overall confidence measure was averaged for all responses.

We had 126 submissions from 20 participants only for this entity disambiguation group. Once we group these submissions by their TaskIDs as shown in Fig. 5.8, a strong co-relation cane be observed between the precision and the confidence values. The lowest precision comes from the first task (TaskID = 1) in which the crowdworkers were asked if the "E. Rodney Canfield" entity in DBLP dataset actually represents the same person as the "Rod Canfield" entity also in DBLP. Almost half of the responses were not correct. This is possibly because both entities were lacking other relations like university affiliations, co-workers, and publications while most of other tasks did have those relations in the original data provided to the workers. On the other hand, participants provided answers with the precision of 0.757 for a task, which was about resolving ambiguity between "Krys J. Kochut" entity in ACM, and "Andrzej Kochut" in DBLP. We believe that a reason for this high precision value is that the relations for both entities were provided, thus it was easier to determine whether the two entities were representing the same person or not.

Another analysis is conducted by grouping the the data according to WorkerIDs as shown in Fig. 5.9. It is interesting to note that almost 70% of the workers have precision values equal to or more than 70%, which is very promising indeed. Once

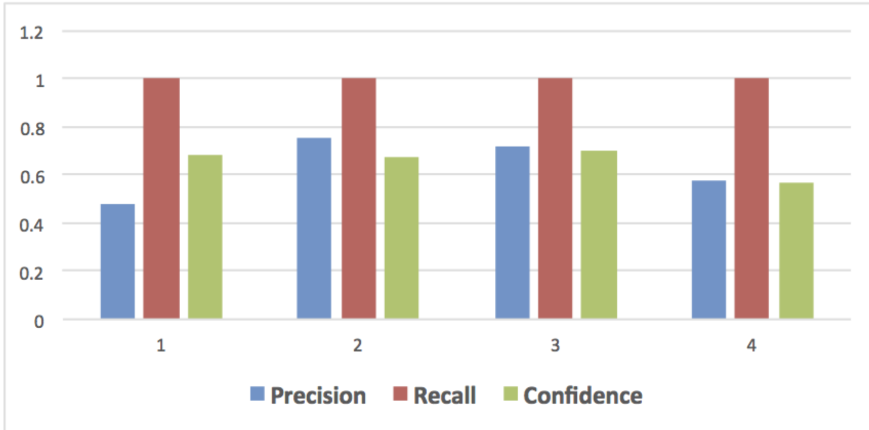


Figure 5.8: Confidence based Analysis for Entity Disambiguation Tasks (Grouped by TaskID)

again, the overall co-relation is obvious between the task precision and the worker confidence measure. A deeper analysis into the submissions of those workers with lower precision indicates one of the two conclusions: (1) the number of submissions by these workers is considerably lower than those of others, and (2) the confidence level indicated by these workers is not high. Also the submissions with the low precisions tend to belong to those cases such as "Walter D. Potter" in ACM & "Walter Potter" in DBLP, where additional relations for those entities are not provided to the workers. As mentioned before, the workers were therefore possibly not able to find out whether those entities represent the same person or not. This is indicated by the correlation between the precision and the confidence measure. In most cases, the confidence measure is indicative of the precision. However, some workers (such as 2, 15 and 19) have lower precision despite indicating high

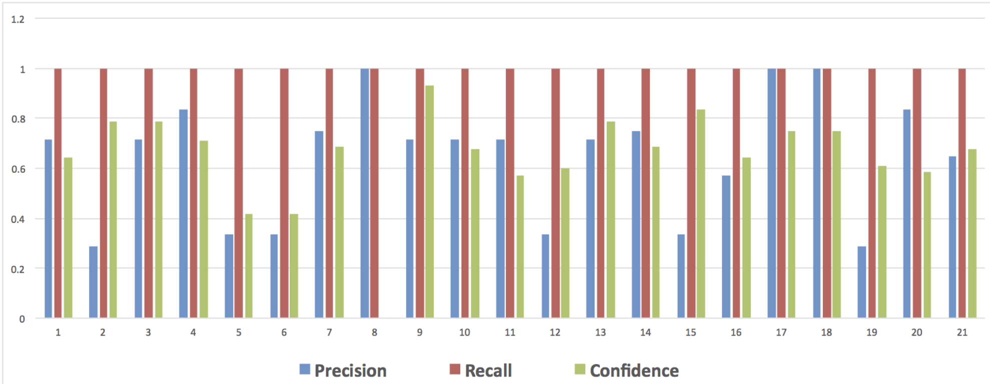


Figure 5.9: Confidence based analysis for Entity Disambiguation tasks (Grouped by WorkerID)

Table 5.6: Candidate Entities from ACM and DBLP for Entity Disambiguation.

Candidate Entity 1(ACM)	Candidate Entity 2(DBLP)
John A. Miller	John Miller
Krys J. Kochut	Andrzej Kochut
Walter D. Potter	Walter Potter
Amit P. Sheth	Amit Sheth
Lakshmi Ramaswamy	Lakshmish Ramaswamy
E. Rodney Canfield	Rod Canfield
Michael A. Covington	Michael Covington
Hamid R. Arabnia	H. R. Arabnia

confidence.

5.3.5 Discussion (Outcome and Results)

Through the work presented in this chapter, we have achieved the design and development of a fully functional semantically enriched task management and workflow generation engine geared to provide LOD management tasks through crowdsourcing using the AMT. We have accomplished the provision for the automatic definition of tasks and the corresponding publishing of HITs using the AMT web service API, their interdependencies and achieving their workflow automation and integration. At the heart of the architecture are ontological specifications and vocabularies (i.e., task profiles) to enable the definition of these workflows. In order to conduct proof-of-concept testing, sample workflows from different data sets in LOD are used; US Census data, and Linked Movie Database are used to validate the functionality of the task engine. Different data access formats are supported such as either file or a live SPARQL endpoint. The generation of a completely generic workflow is validated, which can support any number of SPARQL variables, and any number of HITs. The workflow engine is also suitable to design human-machine cooperative workflows for tasks such as entity disambiguation.

An important question is if the sheer number of RDF triples in LOD can be all reviewed and verified by crowdworkers. Obviously this will require a huge investment in terms of human effort. A more reasonable approach will be using a hybrid approach, which utilizes a combination of Machine and Human Intelligent Tasks in verification workflows. Only mission-critical links that are below a predetermined confidence threshold need to be scrutinized by the crowdworkers. What constitutes a missioncritical link and what is a good confidence threshold are still open research problems. For example, we consider *owl:sameAs* relations, which cross-link multiple data sets as mission- critical links in LOD. We plan to conduct more experimentation using this hybrid approach in the future.

Furthermore, certain verification tasks can be more suitable for human judgement. Typically links that originate from multitude of complex information can be better verified using human intellect. This is partly due to inability of automated information fusion techniques to gather relevant information from disparate information sources and glean new knowledge. Consider establishing new *owl:sameAs* links between researchers in ACM and DBLP data sets. This type of entity disambiguation task not only requires verifying surrounding entities such as co-authors, publications etc. of candidate entities in LOD, but also their affiliations, friends etc. from other web-based resources. This type of complex verification task seems to be more suitable for crowdworkers. However, more simple verification tasks such as verifying number of public schools in a particular county can be easily handled by other tools, which can employ rather simple data extraction techniques. Note that we still use some of these simple tasks in our evaluation, since our research focus is not on distinguishing tasks that are more suitable for crowdsourcing in this paper.

High-value (i.e., mission-critical) links, which can be inspected by the CrowdLink also exist in biomedical ontologies. For example, LinkedLifeData is a platform for semantic data integration of many biomedical sources, including UniProt, PubMed, and EntrezGene. The platform uses an extension of the RDF model that is able to track the provenance of each individual fact in the repository and update the information. It is also possible to execute SPARQL queries on these RDF resources. CrowdLink can be used in verification of LinkedLifeData's mission critical links, many of which use UniProt accession/entry identifiers. These identifiers are also used to cross-link various biomedical data sets in LinkedLifeData, and LOD. A possible experimentation can focus on identification of "bad-links" between the mentioned biomedical ontologies.

VoID is an RDF Schema vocabulary for expressing metadata about RDF data sets. One type of such metadata is the information about links among different RDF data sets, such as the ones in LOD [10]. In many instances, a SPARQL end-point for querying VoID is also provided. It can facilitate querying LOD as one big data set instead of using individual SPARQL end-points. CrowdLink can also use VoID descriptions for link verification, and also contribute in terms of crowd-based link creation using voID.

5.4 Comparison with Related Works

As stated earlier, crowdsourcing has been garnering increasing attention in the research community. A survey on the literature on crowdsourcing, which is categorized according to their applications, algorithms, performances and datasets, can be found in [214]. Another survey focusing on the challenges on how to recruit workers, what they can do, how to combine their contributions, and how to manage abuse is presented in [57]. There have been a number of other studies that look at various performance metrics and statistics on crowdsourcing systems [92]. Some recent research prototypes use crowdsourcing for data management and query processing [120, 170]. For example, Berkeley's CrowdDB tries to answer SQL queries using AMT [68]. Qurk also uses AMT for relational query processing and presents various query execution and optimization challenges, and discusses potential solutions [129]. The sCOOP project views a crowdsourcing service as another database, where facts are computed by human processors [151]. It treats this service at the same level as extensional data, and processes a declarative query that combines information from both.

A system more close to ours facilitates query processing in LOD [175]. However, as mentioned before, our system differs from these systems through the use of semantics, and through the use of well-defined workflows for an improved human computation experience. Task management and specification as well as task interdependencies are not well accounted for in the existing crowdsourcing systems. Moreover, there is no provision for workflow modeling and automation in these systems. One of the most distinctive features of our system is that it provides an adaptive mechanism for LOD management tasks through dynamic provisioning for task dependencies and dependency workflows. On a related note, [48] emphasizes the idea of crowdservicing and inclusion of machine-tasks in addition to human-tasks. His idea shows some parallels with our work on human-machine workflows.

In terms of infrastructure, AMT [1] is among the currently leading platforms for implementing crowdsourcing-based applications on the web. It provides the infrastructure, connectivity and payment mechanisms that enable crowdworkers to perform paid work to answer queries. Although we have used AMT to test our workflow processing, it is worth mentioning that AMT has some limitations. AMT does not address human-machine collaboration; only human subjects can be employed for query processing. In addition, AMT only allows simple task descriptions through its API. For example, one cannot describe a complex crowdsourcing workflow into the system. Although we only mention deficiencies of AMT due to space limitations, other similar systems such as CrowdFlower and Microtask also suffer from the same or similar limitations [48, 214]. [11] also summarizes the limitations of current crowdsourcing platforms. Furthermore, limited existing work investigates on how to support workers to select tasks on crowdsourcing platforms easily and effectively. One approach utilizes the past task preference and performance of a worker to produce a list of available tasks in the order of best matching with the worker during his task selection stage [215].

Amongst the most recent works in crowdsourcing, ZenCrowd identifies entities from natural language text and automatically connects them to the LOD cloud [50]. It also uses crowdsourcing to improve the quality of the links by dynamically generating micro-tasks on AMT. In a similar study, CrowdER studies the problem of entity resolution using crowdsourcing. It describes how machine-only approaches fall short on quality, while crowdsourcing-only approaches are too slow and expensive. Thus, it proposes to use a hybrid human-machine workflow on AMT to address this problem [207]. In another similar research to ours, CROWDMAP proposes a model to acquire human contributions via microtask crowdsourcing for ontology alignment [163].

In a parallel stream, biocuration, the activity of organizing, representing and making biological information accessible to both humans and computers, has become an essential part of biological discovery and biomedical research [96]. Extracting, tagging with controlled vocabularies, and representing data from the literature, are some of the most important and time-consuming tasks in biocuration. Typically, biocurators read the full text of articles and transfer essential information reported in them into a database. Our project can contribute to the biocuration efforts since LOD contains various biological ontologies, and our approach can help in the generation of new knowledge in terms of RDF or OWL triples from biological literature, such as PubMed documents.

5.5 Beyond CrowdLink: Using CrowdSourcing in the LOD

5.5.1 Crowdsourcing and Query Processing

In the context of CrowdLink project, we envision to study beyond link (i.e., triple) creation and verification aspects and investigate querying LOD using crowdsourcing. In this regard, we plan to extend our approach to support real-time querying on the LOD, where measuring response time will be relevant. In this way, users will be able to compose SPARQL queries on the LOD, and receive responses composed of both LOD data sets as well as answers from the crowdworkers. For this purpose, the term CrowdSPARQL can be coined. In a complimentary approach, some parts of LOD can be designated as crowd triples, or parts of triples (i.e., subject, object or predicate). In this case, only portions of the queries that deal with the crowd triples or the triple parts will require crowdsourcing, since either they require human computation or may not be readily available in LOD. Furthermore, crowd querying is typically done using a natural language question-answering.

This may require development of a mechanism for mapping SPARQL queries to natural language queries and vice-versa.

5.5.2 Crowdsourcing and Text-mining

Many tasks involving text-mining are still best accomplished by humans, and we believe that crowdsourcing can leverage collective human abilities in such tasks. As an example, let us consider text-mining tasks in bio-medical literature. Researchers in this area enjoy access to full-text of articles published in numerous journals with on-line access. PubMed Central, created by the U.S. National Institutes of Health's National Library of Medicine, currently offers over 3.1 million full-text scientific articles, all free of charge. This vast volume of biomedical knowledge spurred rapid progress in text mining activities, including extraction of entities and events, as well as relationships among them. For example, text mining has been used to extract information about genes and their synonyms, proteins, mutations, protein-protein interactions, discovering associations between diseases and genes, drug discovery, or discovering pathway networks directly from literature. In all of these text-mining tasks, the best systems have been able to achieve moderate level of accuracy in terms of precision and recall. Nevertheless, processing millions of full-text articles may extract a vast amount of data potentially of interest to biomedical researchers. However, scientists typically access information from well-established and trusted sources, such as GenBank [33], UniProt [12], Reactome[106], or Diseasome[74] and rely on the data accuracy. Consequently, to include automatically extracted data in such databases would require a time-consuming curation process (many biomedical databases curate their data). Crowdsourcing could speed-up the curation of data automatically extracted by various text-mining tasks. Of course, such curation activities would require biomedical experts, and if highest-level researchers would not be readily available, we could rely on a large number of well-trained graduate students. The benefits of a text-mining system coupled with crowdsourcing platform would be felt immediately.

5.6 Conclusions

In conclusion, our work presents a contribution for the linked data management using crowdsourcing. Our system sets apart from other existing systems primarily through the use of semantics based tasks and workflows for an improved human-computation experience. We demonstrate the usefulness of our approach by providing means for automatically generating and managing task workflows pertaining to various LOD link addition, verification and entity disambiguation tasks in the AMT. Experiments reveal that our volunteer workers perform the assigned tasks with high accuracy. We aim to extend our work with more rigorous experimentation and evaluation using a larger number of (potentially paid) crowdworkers and tasks for LOD creation and verification.

Chapter 6

A Semantic Framework for Thematic Annotations of the Qur'an Leveraging Crowds and Experts

وَمَا خَلَقْتُ ٱلْجِنَّ وَٱلْإِنسَ إِلَّا لِيَعْبُدُونِ

And I did not create the jinn and mankind except to worship Me.

— *[Al-Quran, Adh-Dhaariaat,54:17]*

Chapter Overview[1]

In this chapter, we investigate an effective methodology that utilizes the right balance of human and machine computation. We augment traditional ontology engineering approaches with human contribution through established crowdsourcing and human computation methods. We have designed and developed a semantics driven framework, for obtaining thematic annotations from historical and religious knowledge sources. Our framework proposes the use of expertise driven task and crowd profiles for crowdsourcing ontology engineering tasks at varying levels of granularity and knowledge intensiveness. A semantics based task specification model allows for composing automated knowledge acquisition workflows that combine machine computation with crowdsourced annotations at two levels. At the lower level, simple and less knowledge intensive tasks are crowdsourced using the Amazon Mechanical Turk platform. At the higher level for more knowledge intensive tasks, skilled workers and experts are engaged through a custom web application. We demonstrate the effectiveness of this hybrid model for several key knowledge engineering tasks such as thematic disambiguation and thematic annotation, in the Qur'an.

[1]This chapter is partially published as:
Basharat, A., Rasheed, K., Arpinar, I.B., "Harnessing Crowds and Experts for Semantic Annotation of the Qur'an", *Proceedings of the ISWC 2016 Posters & Demonstrations Track co-located with 15th International Semantic Web Conference (ISWC 2016). CEUR-WP, Vol-1690, p96.*
and
Basharat, A., I. Budak Arpinar, Rasheed, Khaled, "Leveraging Crowdsourcing for the Thematic Annotation of the Qur'an." *In Proceedings of the 25th International Conference Companion on World Wide Web*, pp. 13-14. International World Wide Web Conferences Steering Committee, 2016.

6.1 Introduction

Efforts towards semantic annotation and knowledge engineering in specialized and knowledge intensive domains continue to present challenges to the semantic web researchers. The Qur'an is one of the most widely read and studied books. Its original script is in the Arabic language, which is rich in both its morphology and semantics.

Thematic annotation of religious texts, in particular, the classical sources of knowledge in the Islamic domain in the Arabic language, has not received much attention, partly owing to time and knowledge constraints required for such an annotation process.

As part of this research, we undertook the task of thematic disambiguation and annotation (as part of formal and standardized knowledge modelling and ontology engineering activities) of the Qur'anic verses by augmenting the traditional information extraction and text mining techniques with crowdsourcing methods. The need for crowdsourcing stems from the fact that for the Qur'anic knowledge, a high level of accuracy and reliability is desired, given the sensitivity of the knowledge at hand. Pure computational approaches fail to meet this standard and the contribution from human experts is indispensable. However, finding experts is greatly time consuming and this makes the process of obtaining semantic annotations non-scalable. Motivated by the success of human-computation and crowdsourcing methods, we therefore attempt to investigate the usefulness of these approaches for knowledge intensive and specialized domains, such as the one encompassed by the Qur'an.

The novel aspect of this research is that, we augment the traditional model of crowdsourcing with the application of specialized human computation methods such as nichesourcing ([49], [149]) in an attempt to scale this process of annotation while maintaining reliability and ensuring quality control and validation. Nichesourcing or expertsourcing extends the idea of engaging skilled and knowledgeable persons in place of faceless crowds for human driven tasks. We employ nichesourcing as means of augmenting traditional crowdsourcing methods rather than as an alternate.

In this chapter, we present our methodology for the implementation of hybrid approach engaging crowds and experts and present the results of our study focussing on two knowledge intensive tasks: the thematic disambiguation and annotation of Qur'anic verses using its Arabic script.

We determined that several tasks are rather knowledge intensive and require domain expertise. In this case, the knowledge of Qur'anic Arabic is considered imperative, since the annotation of the Arabic verses require understanding the context of a given Arabic word in the verse.

6.1.1 Research Questions

Emerging semantic web research, realizing the importance of human intelligence and contribution, has been leveraging various models of human computation and crowdsourcing to effectively overcome the inherent knowledge acquisition bottleneck in large scale engineering of semantic knowledge. However, an effective crowdsourcing methodology that utilizes the right balance of human and machine computation, while engaging the appropriate crowd especially for knowledge intensive tasks in specialized domains is still a matter that demands further insights and investigation. To address this, we have designed and developed a semantics driven framework for obtaining thematic annotations from historical and religious knowledge sources.

The proposed framework will be validated through contextualized application to the domain of Islamic and Religious texts, where the overall vision is to provide means to enable efficient and reliable knowledge discovery for religious knowledge seekers for a more meaningful knowledge seeking and learning experience. Some of the key research questions that will be addressed as part of the process are as follows:

RQ-1: What is the amenability of crowdsourcing ontology engineering tasks in knowledge intensive domains? Can the tasks of thematic disambiguation, semantic annotation and thematic interlinking be reliably crowdsourced in the specialized and knowledge intensive domains such as the Islamic knowledge?

RQ-2: Is there significant performance gap between experts and crowd workers when performing such tasks?

RQ-3: Can the contributions from crowd workers and experts be reliably and efficiently combined for the purpose of knowledge engineering, using semantics driven, hybrid and iterative workflows? What methodological considerations would enable such an effective synergy?

6.1.2 Contributions

Our work is the first of its kind that leverages semantic annotations for generating an iterative workflow of machine computation, crowd workers and experts for the purpose of Ontology Engineering tasks.

We introduce the conceptual methodology for generating task and workflow specifications, in the form of semantic annotations for a given ontology schema. This is done for two tasks namely thematic disambiguation and thematic annotation in the Qur'an.

We also present the design and implementation of a hybrid framework to harness crowd and expert contributions illustrating the execution of iterative workflows.

We evaluate our methodology and framework based on application to two task types: namely thematic disambiguation and thematic annotation in the Qur'an.

6.2 Problem Definition: Thematic Annotation in the Qur'an

As part of this research, we aimed to augment the available thematic hierarchies of the Qur'an to include annotations for the various themes at sub-verse level, a level deeper than what existing thematic hierarchies provide. The motivation for this comes from the diversity of thematic coverage that Qur'anic verses provide at an individual and collective level. The Qur'anic verses span different lengths; while some verses may be as short as a word or few words, others may span half a page or an entire page of a standard sized book. This inspires the need for increasing the level of granularity at which the *thematic assertions* for each verse are classified. We take our initial hierarchy from QuranyTopics datasource[2], which contains a hierarchy of themes, hand-crafted from a classical source. This is one of the only data sources, that provides an authentic, concept driven, thematic classification of the Qur'anic verses. However, the thematic classification in this resource is not only limited only to the verse level, its coverage of the concepts is not exhaustive.

As part of this research, not only do we propose sub-verse level annotation, we also make the distinction between *explicit* and *implicit assertions* of a *theme* as shown by a segment of the ontology schema designed for obtaining thematic annotations in Figure 6.1. Explicit assertions are amenable to be obtained through text mining and NLP techniques, such as, the direct occurrence of the word or a phrase that directly indicates the *manifestation* of a particular theme. However, even when a word or a phrase explicitly appears in a verse, in the Arabic language, it may manifest different meanings in different contextual settings. When modelling such thematic assertions, at sub-verse level, disambiguation by a human expert becomes indispensable. To add to the challenge, themes may also exhibit implicitly; whereby, the same theme may be manifested using not just a mere difference of word morphology, rather, a difference in expression or rhetoric. It is extremely difficult to extract such implicit thematic assertions via automated techniques, therefore, it becomes imperative that human contribution be sought.

Take for an example the concept of God. In the Qur'an, various names are

[2]http://quranytopics.appspot.com

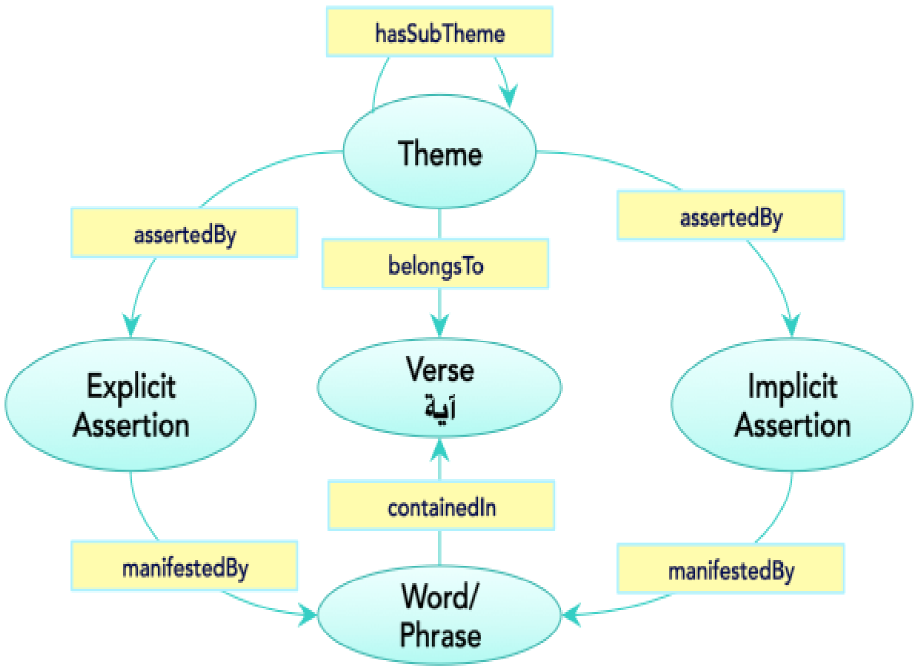


Figure 6.1: A segment of the thematic annotation knowledge model populated via the crowd

given to God (referred to as Allah in the Qur'an), based on His attributes and qualities. One such name given to him is Al-Baseer (The all seeing). This phrase explicitly occurs in the Qur'an several times, however, in each occurrence it may or may not refer to Allah. Therefore, it becomes imperative, that even the seemingly explicit assertions of a particular theme be disambiguated as positive or negative occurrences and be annotated accordingly.

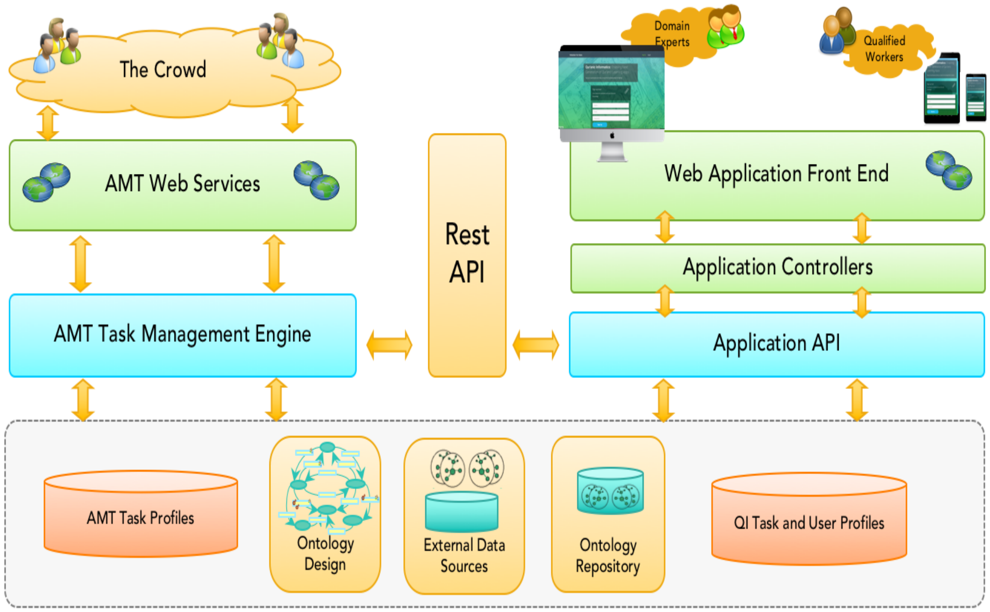


Figure 6.2: Hybrid Architecture for Harnessing Crowd and Expert Annotations

6.3 Hybrid Architecture for Harnessing Crowd and Expert Annotations

We designed and implemented a hybrid workflow architecture that connects a crowdsourcing framework with an expertsourcing application as shown in the Figure 6.2.

6.3.1 Crowdsourcing Stage

We design a task management engine that is responsible for generating tasks, retrieving and aggregating results. The tasks are published on the Amazon Mechanical Turk (AMT)[3] platform. A complete workflow management system is implemented (a derivative of a workflow model for Linked Data Management presented in [28]), which includes means for generating dynamic tasks from a range of task profiles. An ontology schema such as the one given in Figure 6.1 guides the semantic annotation process. This serves as input to the *Task Generation and Design* stage, which creates an annotation or disambiguation relevant task to be crowdsourced based on the nature of the entity, relation or both, as specified in the *task specification*. The relevant task input is generated by retrieving relevant candidate verses from the available data sources such as the Semantic Qur'an [168] dataset or the quranontology[4].

6.3.2 Task and Workflow Design

We introduce two key workflows and usecases in the context of which the system has been developed and tested. The AMT crowd performs the *thematic disambiguation and annotation tasks*. Both tasks are based on the Arabic script of the Qur'an, therefore, requiring the crowd workers to be familiar with the Arabic language of the Qur'an.

Thematic Disambiguation Task

In the Arabic language, there may be many words or their derivatives, that originate from a different 'root', yet may exhibit the same meaning. It is also possible for the same word to have contrastingly different meanings depending on the con-

[3]http://www.mturk.com
[4]http://quranontology.com

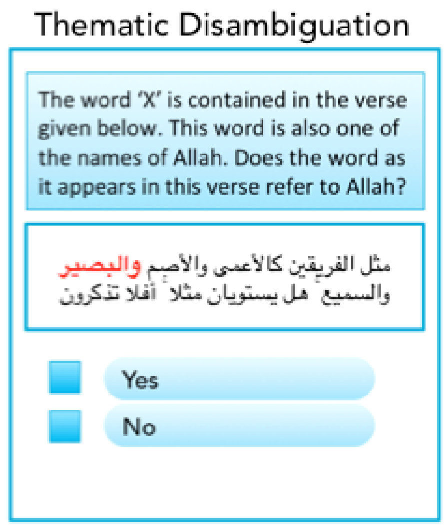


Figure 6.3: An Example of Learner tasks introduced in the learning scenarios (Thematic Disambiguation)

text it occurs in. Therefore, given that an existence of a particular phrase or a word in a verse, it needs to be disambiguated whether the word refers to the presented theme or not.

A sample task of this nature is shown in Figure 6.3.

Figure 6.4 shows what a typical disambiguation task workflow in the domain looks like.

For the disambiguation task, a question is presented to the crowd, which includes a verse, along with a highlighted, candidate explicit assertion for the given theme, and the crowd responds by declaring this assertion as either a positive or negative by determining if the occurrence is a true occurrence of the given theme.

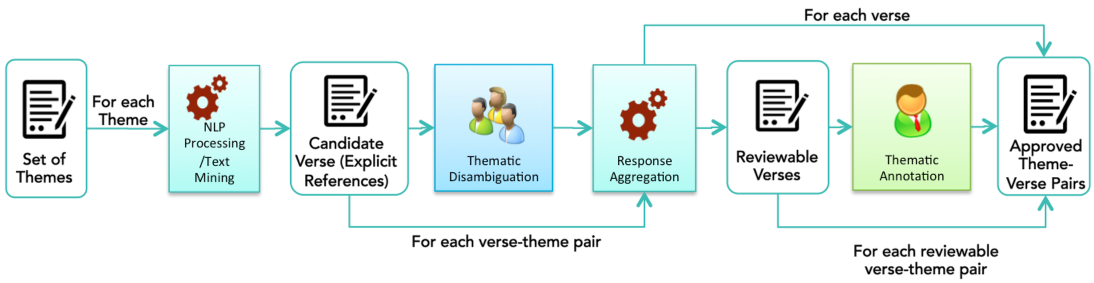


Figure 6.4: Human-Machine Computation Workflow for Disambiguation Task

Thematic Annotation Task

This task is used to annotate portions of a Verse or a narration that may implicitly refer to a given theme. Figure 6.6 shows what a typical annotation task workflow in the domain looks like. The annotation tasks require deeper knowledge and understanding of the Arabic text. The crowd determines whether the given verse contains any implicit reference to the given theme. If their response is positive, then they are also required to provide the portion of the verse (a meaningful phrase or a word) that implies the presence of the theme. A sample template for this is shown in Figure 6.5.

6.3.3 Expertsourcing Stage

For this purpose we designed a custom web application to engage with experts. A RestAPI connects the crowdsourcing task management engine with the expert-sourcing application. The tasks are sent to the remote application and experts are notified when the tasks become available. The experts also see the candidate

Figure 6.5: An Example of Learner task (Thematic Annotation)

responses collected from the crowdsourcing stage. The experts have either the option to choose from the available annotations (collected during the crowdsourcing stage) or provide their own if they do not agree with either one. An example is shown in Fig 6.7. We present the same task to three experts to analyze the annotation agreements. The approved and validated annotations are passed on for *Ontology Population* and linked with existing data sources.

6.3.4 Decision Analytics

We collect and aggregate the responses based on statistical measures of aggregation. Weighted confidence measures and thresholds are applied. Based on this aggregation, the completed tasks are marked as either *Approved* or *Reviewable*.

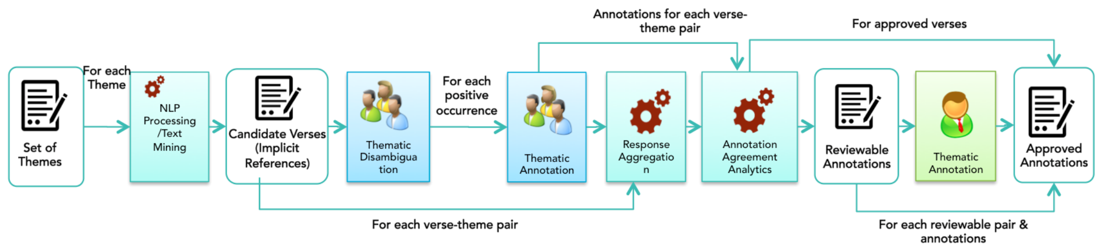


Figure 6.6: Human-Machine Computation Workflow for Annotation Task

A high confidence and aggregation threshold is applied for the approved tasks. This process of decision analytics results in identifying the candidate tasks for the expertsourcing stage. The tasks marked as reviewable, which fail to meet the agreement thresholds, are sent off for expert annotations. Following are the mechanisms employed in the decision analytics stage:

Agreement Analytics for the Crowdsourcing Stage

The Figure 6.8 shows the decision analytics process implemented for approving the tasks that reach agreement based on the crowd submissions. The table 6.1 shows the notation used in process.

Each task may be performed by upto 5 crowd workers. The Inter-Annotator Agreement is calculated using the decision analytic algorithm illustrated in Figure 6.8. N_A denotes the number of annotations or task submissions obtained for a given task. If this number is more than the *minimum response threshold*, T_{MIN}, then the *response agreement*, C_{RA}, is calculated for each possible response for the

Table 6.1: Notation used in the Decision Analytics Process

Notation	**Meaning**
T_{RA}	Response Agreement Threshold
T_{AA}	Annotation Agreement Threshold
T_{MIN}	Min Response Threshold
T_C	Confidence Threshold
T_{UA}	Unique Annotation Threshold
$RA - Approved$	Task approved by Response Agreement
$AA - Approved$	Task approved by both response and annotation agreement
$nRA - Reviewable$	Task marked as Reviewable with no Response Agreement
$nAA - Reviewable$	Task reached response agreement but not annotation agreement
$LC - Reviewable$	Task marked as Reviewable with response agreement but low confidence
$MA - Reviewable$	Task marked as Reviewable with response agreement and annotation agreement but with more than expected unique annotations

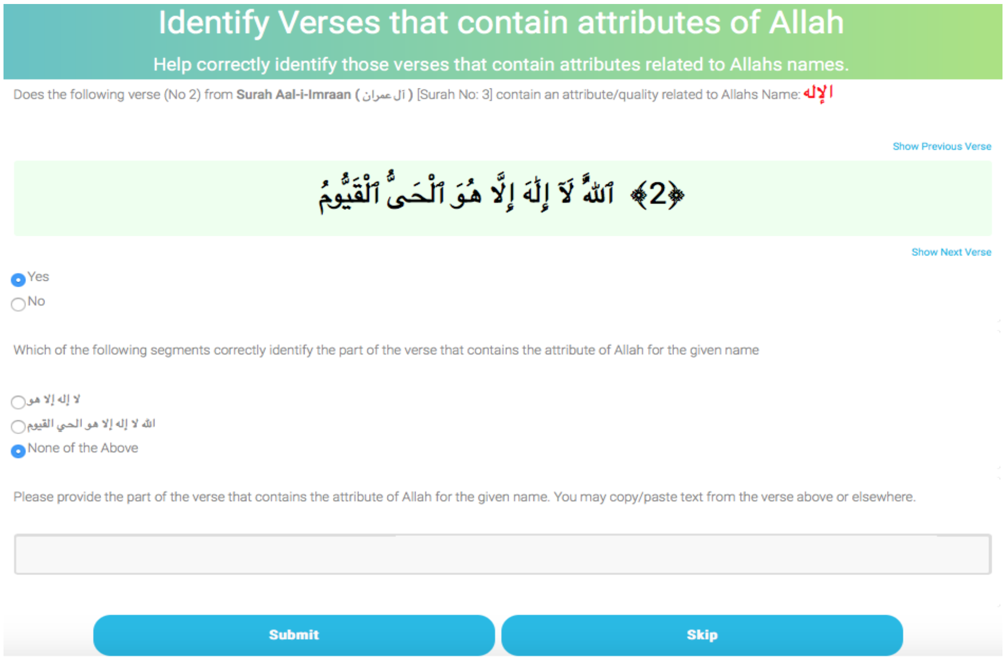


Figure 6.7: Task Design for Thematic Annotation of Qur'anic verses

given task.

$$C_{RA} = \sum_{i=1}^{m} R_i \tag{6.1}$$

where R_i denotes the i^{th} response for the given task. So for the two task types, *disambiguation* and *annotation*, there are two possible responses *Yes* or *No*.C_{RA} is calculated for each of these responses. If either of these exceeds the *response agreement threshold*, T_{RA}, then the *response confidence measure*, C_{RC} will be calculated. If none of C_{RA} meets the T_{RA}, it implies that the desired agreement has not been reached and the task is marked as *Reviewable*.

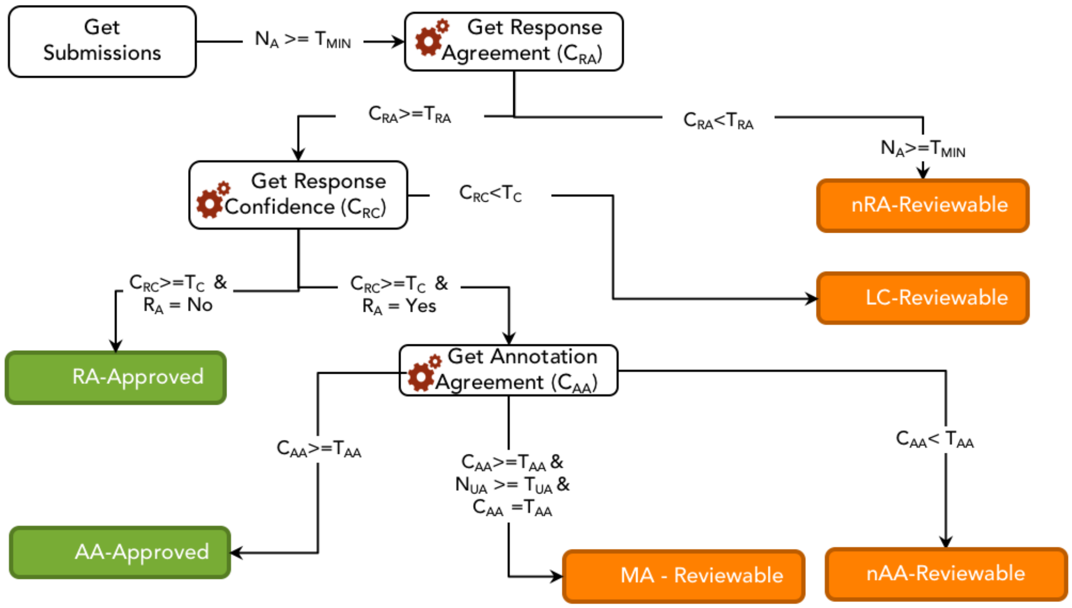


Figure 6.8: Decision Model used for Approving the Tasks based on Crowd Submissions and marking the ones to be sent to the Experts as Reviewable

The C_{RC} is calculated as follows:

$$C_{RC} = \frac{\sum_{i=1}^{m} S(C_i)}{m} \tag{6.2}$$

Equation 6.2 gives the average confidence as indicated by the crowdworkers in the responses that they provide. $S(C_i)$ represents the function that assigns a numeric scale to the confidence category selected by the worker. For instance $S(VHigh)$ is assigned a numeric value 5 and so on.

If the C_{RC} is less than the *confidence threshold*, T_C, then the task is marked as *Reviewable*.

In the case of *disambiguation task*, the analytic cycle ends here. The task is marked *Approved* if the C_{RC} meets the T_C.

In the case of *annotation task*, the same happens if the *agreed response*,R_A is *No*. However, if the R_A, then it implies that an annotation must be provided by the crowd worker and the *inter-annotation agreement*, C_{AA} must be determined.

This is calculated by creating a mxm matrix, M, whereby, m is the number of annotations obtained.

The similarity between two annotation vectors, a_i and a_j, denoted as, $M_{i,j}(\bar{a}_i, \bar{a}_j)$, is measured using the Jaccard coefficient. The Jaccard coefficient measures similarity as the intersection divided by the union of the entities. For the annotation vectors, the Jaccard coefficient computes the ratio between the dot product and the sum of the squared norms minus the dot product of the given annotation vectors. The definition is given in Equation (6.3), similar to the Equation 4.5.

$$M_{i,j}(\bar{a}_i, \bar{a}_j) = \frac{\bar{a}_i \cdot \bar{a}_j}{|\bar{a}_i|^2 + |\bar{a}_j|^2 - \bar{a}_i \cdot \bar{a}_j} \tag{6.3}$$

The Jaccard coefficient is a similarity measure and ranges between 0 and 1. It is 1 when the $a_i = a_j$ and 0 when a_i and a_j are disjoint, where 1 means the annotations are the same and 0 means the annotations are completely different.

$$C_{AA} = Max(\sum_{i=1}^{m} M_{i,j}) \tag{6.4}$$

If the C_{AA} meets the *annotation agreement threshold*, T_{AA}, then the task is marked as *Approved*, else it is marked as *Reviewable*.

Role of the Measure of Confidence

As a form of a quality measure, the crowd is also required to provide a confidence level (ranging from Very High to Very low), to indicate their confidence in their response. In other words, this reflects workers' own confidence in the correctness of the response they provided. A task will only be approved if the average confidence measure of the aggregated responses exceed the confidence threshold. The task is marked for review otherwise and will be sent to the expert.

Agreement Analytics for the Expertsourcing Stage

The process of agreement analytics in the expert sourcing stage is similar to the one in the crowdsourcing stage. Each reviewable task may be performed by up to 3 experts. Once the three submissions are received, the task status is marked complete and the agreement analytics are performed similar to the one described for the crowdsourcing stage. In the case of experts, the confidence measure is not sought. The agreement and annotation thresholds are also different as explained below.

6.3.5 Agreement and Confidence Thresholds

Table 6.2 shows the thresholds used for establishing agreement analytics in the crowdsourcing and the expert sourcing stages. Note that the thresholds depend on the task at hand.

For the disambiguation task, the response agreement threshold is selected as 4, whereas the confidence threshold is 3. The confidence is the average confidence for the majority response. If we assume that the probability of someone accidentally giving a wrong response is 25%, then with 5 maximum responses and an agreement threshold of 4, we ensure at least 96% accuracy.

For the annotation task, the annotation agreement threshold is selected as 3. The reason for the lower threshold is that we take annotation agreement with 0 distance. And if the number of unique annotations reaches 3 or more, we mark the task as reviewable even if the agreement is reached.

For the expert tasks, since there are only 3 submissions, we choose 3 as the response agreement threshold for the disambiguation tasks.

For the annotation task, we allow 2 as the threshold. This is done because we assume that the probability of an expert giving a wrong response would be 10% unlike that of the crowd. We analyze the impact of these thresholds in the results and analysis section.

Table 6.2: Agreement and Confidence Thresholds for Crowd and Expert Output Decision Analytics

Contributor	**Task Type**	**No of Responses**	**Threshold**	**Confidence**
Crowd	Disambiguation	5	4	3
Crowd	Disambiguation	4	4	3
Crowd	Annotation	5	3	NA
Expert	Disambiguation	3	3	NA
Expert	Annotation	3	2	NA

Table 6.3: Results for Disambiguation and Annotation Tasks for both stages

	Crowd Tasks		**Expert Tasks**	
Task	Disambiguation	Annotation	Disambiguation	Annotation
Approved	1267	477	34	96
Reviewable	40	107	6	11
Total	1307	584	40	107

6.4 Experimental Evaluation

6.4.1 Details of Experimentation

MTurk Setup

The experimental setup assigned each task to 5 crowd workers. Any task with less than 5 submissions was ignored.

Expertsourcing Setup

For the reviewable crowd tasks that were sent to the experts, 3 experts were assigned to each reviewable task. Up to 7 experts were invited to participate in the task review process. The system for this phase worked on an invite-only basis.

6.4.2 Results and Discussion

Table 6.3 shows the results obtained. We only included those tasks in the results that reached the required response threshold. Any tasks that obtained fewer than desired threshold were marked as incomplete and were not included in the results.

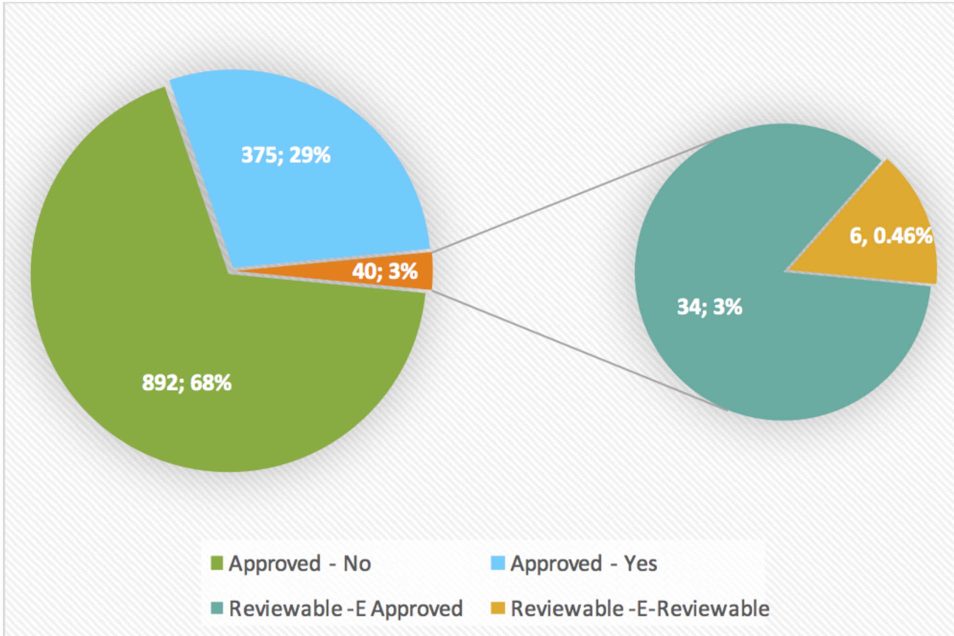


Figure 6.9: Distribution of Number and percentage of Disambiguation Task from Crowd and Expert Sourcing stage

6.4.3 Results from the Crowdsourcing Stage

Figure 6.9 and Figure 6.10 further illustrate the results obtained in the crowdsourcing and expertsourcing stages of the disambiguation and the annotation tasks respectively.

Results for Disambiguation Task

As was discussed in the Section 6.3.4, the disambiguation task may have two possible responses, a *Yes* to indicate a positive response and a *No* to indicate the negative. As illustrated in the Figure 6.9, 68% tasks were approved with a positive

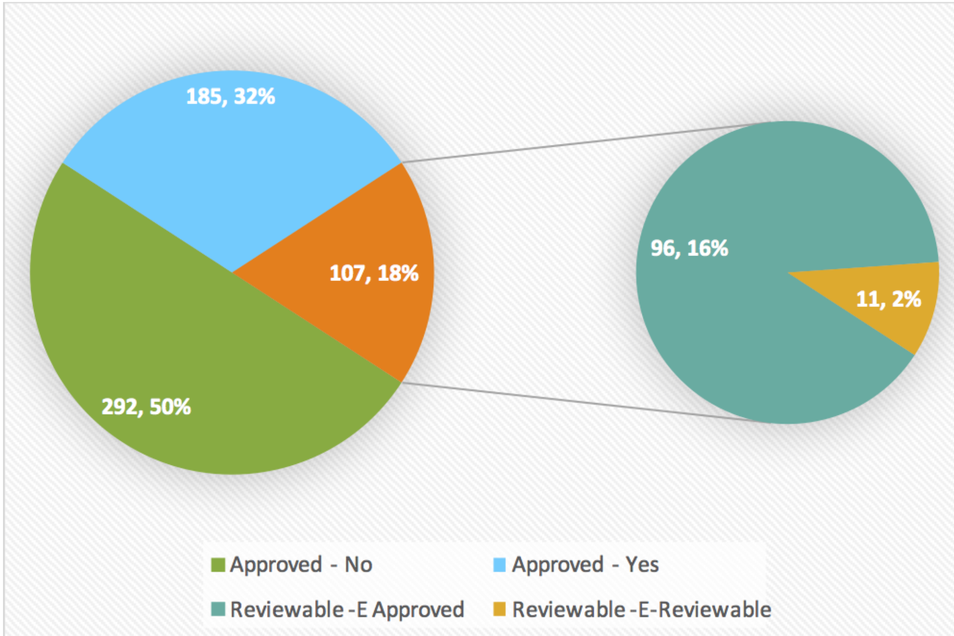


Figure 6.10: Distribution of Number and percentage of Annotation Task from Crowd and Expert Sourcing stage

response, while 29% were approved with a negative response. Only 40 tasks, which accounts for 3% of the total tasks were sent to the experts for review.

Results for the Annotation Task

As was discussed in the Section 6.3.4, the annotation task will have an annotation provided as a response from the given verse after indicating a positive or a negative response. As illustrated in the Figure 6.10, 50% tasks were approved with a negative response, while 32% were approved with a positive response. This also indicates that these 32% tasks contained annotations from the crowd that reached an inter-annotator agreement. 107 tasks, which accounts for about 18% of the total tasks were sent to the experts for review.

6.4.4 Results from the Expertsourcing Stage

Expertsourced Disambiguation Tasks

Of these 40 tasks, there were only 6 tasks that received 1 different response from the majority response.

If 2 out of 3 majority response agreement was taken amongst the experts all 40 tasks would reach an *E-Approved* status i.e. an approved by the experts.

An additional administrative round of evaluation was done for the 6 tasks that did not achieve a consensus amongst the three experts. The majority was infact correct.

Of the reasons that may have led to the disagreement include: 1) Failure to interpret the question, 2) Accidental error.

Expertsourced Annotation Tasks

Of these 107 tasks, there were only 11 tasks that did not reach a well defined agreement even after the expertsourcing stage and needed further administrative or super review.

For the Annotation tasks, a majority threshold was decided as 2 out of 3 to reach an agreement. Any cases that had only two or three distinct annotations were the one's that clearly did not reach an agreement. For such 11 tasks, the annotations were finalized based on super review.

For the tasks which agreed on a negative response but did not have a 3 out of 3 consensus went through a super review for the sake of quality control. Of these, there were 2 tasks which agreed on a false response.

An interesting observation is that the 11 tasks which were super-reviewed, were based on only 4 distinct themes. This indicates that some themes may be more likely to need deeper knowledge and understanding than others.

From the results, it can be concluded that our system is indeed very robust to error and is capable of obtaining high quality annotations. The cases where there is arguably or apparently a disagreement are those consensus cannot be reached even after the expert reviews. These cases indicate that some annotations are a matter of personal taste and judgement and cannot be concretized as others can be. Most of these annotations cannot be classified as wrong, nor better than the others based on an automated choice.

6.4.5 Scoring of Crowd Submissions

One of the important stages following the task analytics stage is to score and mark the original crowd submissions as either *Accepted* or *Rejected*. If a given crowd submission, matches the majority agreed response, it is marked as *RA-Accepted*, else it is marked as *RA-Rejected*. In the case of annotation tasks, there will be additional scoring based on *annotation agreement*. If the annotation in the crowd submission matches the agreed upon annotation (i.e. inter-annotation distance is 0) then the submission is marked as *AA-Accepted*.

For the submissions that make up the reviewable tasks, the scoring of the crowd

Table 6.4: Crowd Submission Scoring for Disambiguation Task

Task State	**Submission State**	**Count**
Approved	RA-Accepted	6134
Approved	RA-Rejected	140
Reviewable	E-RA-Accepted	117
Reviewable	E-RA-Rejected	76

submissions is done based on the agreements obtained from the expertsourcing stage. This is indicated by *E-RA-Accepted* or *E-RA-Rejected* to indicate the status of the response, and *E-AA-Accepted* or *E-AA-Rejected* to indicate match with the agreed annotation.

Table 6.4 show the results for the disambiguation task and the table 6.5 shows the results for the annotation task.

As can be noted, the two combinations *RA-Rejected-AA-Accepted* and *E-RA-Rejected-AA-Accepted* indicates that the response did not match the majority or the expert response, however the annotation did. This is an interesting combination. However, the small number indicates that this may be accidental error.

6.4.6 Measure of Quality

Since we are dealing with tasks, which do not have a pre-established ground truth, determining the quality of the crowdsourced task presents a challenge. The methodology adopted for measuring the quality of the crowdsourced tasks is as follows. From the tasks that were approved, 100 tasks from each category were selected at random and were presented to a qualified expert using the expert-

Table 6.5: Crowd Submission Scoring for Annotation Task

Task State	**Submission State**	**Count**
Approved	RA-Accepted-AA-Accepted	2296
Approved	RA-Accepted-AA-Rejected	85
Approved	RA-Rejected-AA-Accepted	9
Approved	RA-Rejected-AA-Rejected	89
Reviewable	E-RA-Accepted	126
Reviewable	E-RA-Rejected	60
Reviewable	E-RA-Accepted-E-AA-Accepted	155
Reviewable	E-RA-Accepted-E-AA-Rejected	158
Reviewable	E-RA-Rejected-E-AA-Accepted	3
Reviewable	E-RA-Rejected-E-AA-Rejected	25

sourcing application. The expert's response was then matched against the crowd submissions and the agreed response determined by our system. Table 6.6 shows the number of task submissions which are evaluated using the system and expert response. It can be seen that for the disambiguation task, there was 100% agreement between the expert response and the system response. For the annotation task, the disagreement occurred only for one task's five assignments, where the expert selected the annotation which was only slightly different from the one agreed upon by the system.

This is a clear indicator of the robustness and accuracy of the system. It also indicates, that the choice of thresholds for our agreement analytics module was suitable, and the crowd actually meets the competency threshold that is much higher.

Table 6.6: Results showing the number of submissions as evaluated based on the system and the expert responses

	Disambiguation Task		**Annotation Task**	
Submission Status	System Response	Expert Response	System Response	Expert Response
RA-Accepted	484	484	488	488
RA-Rejected	12	12	21	21
AA-Accepted	NA	NA	4	1
AA-Rejected	NA	NA	1	4
Total	496	496	514	514

6.4.7 Comparative Analysis : Expert and Crowd Contributions

Table 6.7 the state of tasks after the decision analytics of both the crowdsourcing and the expertsourcing stage. It is interesting to note that almost 50% of the annotations obtained from experts actually agree with those of the crowd.

This provides us with an important implication. By increasing the number of contributions from the crowd, a reasonable agreement may be reached. This provided the basis and the motivation for the study presented in Chapter 7.

Table 6.7: Comparative Analysis of Crowd and Expert Tasks (Annotation Tasks)

Task State (Type) - Crowd Stage	**Review State (Expert Stage**	**No of Tasks**	**No of Tasks that Agree on same Annotation**
Reviewable	E-Approved	24	0
Reviewable	E-AA-Approved	72	49
Reviewable	E-NAA-Reviewable	11	8
	Total	107	57

6.4.8 Discussion

The results of our exploratory study provide interesting insights into the application of human computation methods to knowledge intensive tasks. Our task design involved the thematic disambiguation and annotation of the Qur'anic verses based on the original Arabic script. For the disambiguation task, 99% tasks were able to reach an agreement by combining contributions of crowds and experts. Only 4% tasks needed expert contributions. For the annotation tasks, 10% tasks did not reach an agreement with both crowd and expert contributions. An investigation into these indicate that closed agreement is not always possible for open ended annotation tasks, and the response aggregation needs to take into account for such tasks. Our knowledge acquisition and review workflow to elicit expert annotations for the tasks that exhibit significant variation in annotations from the crowd contributors presents a promising method. We utilize annotation agreement and distance analytics to route the appropriate tasks that need expert contributions. Our results suggest that such a hybrid approach indeed creates for a more scalable, yet accurate and reliable annotation process whereby appropriate annotation tasks are assigned to different annotator groups according to different skill sets.

6.5 Conclusions

The results of the study presented strongly suggest that the crowd can significantly assist in scaling the knowledge engineering activities such as knowledge formalization, and semantic annotation with reasonable reliability for several disambiguation and annotation tasks. Our efforts have also demonstrated that augmenting the workflow to obtain expert validations for the reviewable tasks (through our own custom web framework) ensures that high quality annotations may be obtained, and ambiguous cases may be resolved. Our study also inspires the investigation of using an adaptive measure of obtaining crowd annotations to reach a substantial agreement, without the need to go to the experts. To this end, we plan to experiment towards achieving the right balance of crowd and expert tasks in later studies to follow (as done in Chapter 7). We plan to increase the size of the study, and experiment with a range of other task designs of varying complexity. The study clearly shows the potential benefit of crowdsourcing, augmented with expertsourcing workflows that can be harnessed by knowledge intensive and expertise driven domains.

Chapter 7

Learnersourcing Thematic and Inter-Contextual Annotations from Islamic Texts

... وَمَن يَتَّقِ ٱللَّهَ يَجْعَل لَّهُۥ مَخْرَجًا
وَيَرْزُقْهُ مِنْ حَيْثُ لَا يَحْتَسِبُ ۚ وَمَن يَتَوَكَّلْ عَلَى ٱللَّهِ فَهُوَ
حَسْبُهُۥٓ ۚ إِنَّ ٱللَّهَ بَٰلِغُ أَمْرِهِۦ ۚ قَدْ جَعَلَ ٱللَّهُ لِكُلِّ شَىْءٍ قَدْرًا

> ... And whoever fears Allah - He will make for him a way out. And will provide for him from where he does not expect. And whoever relies upon Allah - then He is sufficient for him. Indeed, Allah will accomplish His purpose. Allah has already set for everything a [decreed] extent.
>
> — *[Al-Quran, At-Talaaq,65:2-3]*

Chapter Overview[1]

In this chapter, we introduce an approach for obtaining semantic annotations in specialized and knowledge intensive domains. In particular, we consider the case of classical and historic Islamic texts, primarily the Qur'an and the books of Prophetic narrations called the Hadith. We propose formal and scalable knowledge acquisition workflows for thematic classification, annotation and interlinking for these texts; this is done at various levels of granularity that existing research fails to address. We design a composite *'semantics driven learnersourcing'* workflow, which leverages primarily upon students engaging in typical knowledge seeking and learning scenarios, and embeds within them, semantic annotation tasks. Tightly integrated within the workflow is an 'expert-sourcing' component that ensures annotation quality and reliability. Therefore, quantitative measures of ensuring annotation quality are woven into the very fabric of learnersourcing.

7.1 Introduction

In the recent years, there has been growing recent interest in the particular method of learnersourcing [72, 73, 118, 139, 209]. Learnersourcing is seen as a specialized derivative of crowdsourcing [209], whereby learners are incentivized to perform a

[1] Partially appeared in:
Basharat, Amna. "Learnersourcing Thematic and Inter-Contextual Annotations from Islamic Texts." *In Proceedings of the 2016 CHI Conference Extended Abstracts on Human Factors in Computing Systems*, pp. 92-97. ACM, 2016.
Full version to Appear:
Basharat, A., Rasheed, K., Arpinar, I.B., Semantics-Driven Learnersourcing Workflows for Thematic and Inter-Contextual Annotation of Islamic Texts, *In The 35th Annual ACM conference on Human factors in computing systems (SIG-CHI'17).*

task based on their desire to learn. This is unlike traditional crowdsourcing, in which a crowd worker is paid to perform a micro-task. Learnersourcing approaches claim the possibility of attaining higher quality responses than traditional crowdsourcing, given that learners are more intrinsically motivated to learn the given content, and may also benefit from the task itself [73, 209].

Motivated by these proposed benefits of leanersourcing, in this research, we explore how it can be applied to generating thematic classifications, annotations and interlinking of diverse knowledge sources. We apply the learnersourcing methodology in particular context of the Qur'an, which is one of the most widely studied books in the world (Muslims consider it as their Holy Book, a spoken word from God, Allah). The study of the Qur'an is undertaken at various levels from formal to literal. Formal Qur'anic education is based on the centuries old, classical sources of knowledge which primarily include the Qur'an and Prophetic Narrations termed as Hadith, compiled in numerous classical books of Hadith. These sources use the classical Arabic scripts, which require interpretation and deep understanding, even for the native speakers of modern Arabic.

We believe that the domain of Qur'anic studies is ideally suited for benefitting from learnersourcing rather than crowdsourcing given the specific nature of the content. There is always a strong desire to learn these texts by individuals across the world, which we believe can be potentially harnessed. The broader aim of this research is to facilitate the process of both formal and informal learning and research of the Qur'anic texts by semantically modeling and interlinking the various knowledge sources (through formalized knowledge models e.g. ontologies)

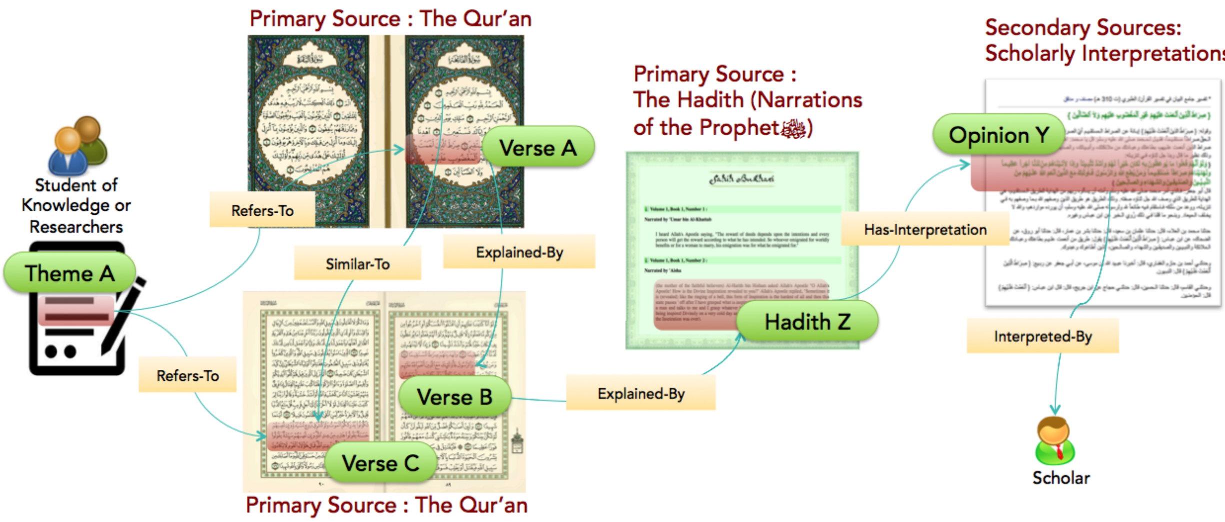


Figure 7.1: A typical path of knowledge synthesis followed by Qur'anic learners

that are involved in developing a deep and thorough understanding of the texts-thus making relevant knowledge sources more discoverable and accessible to the learners.

However, within the realm of (semantic web based) knowledge engineering tasks, such as semantic annotation and data interlinking, several levels of complexity may be encountered. Some tasks are simple, while others are more knowledge intensive. While some may reasonably be amenable to computational approaches, others need domain specific expert annotations and judgements. Therefore, not all tasks are fit for general purpose crowdsourcing. This is specifically true for the domain of islamic knowledge, owing to the specialized nature of the learning needs presented by diverse users, and heterogenous, multilingual knowledge resources. Emerging research also recognizes nichesourcing [49, 149] or expertsourcing [158], as a natural step in the evolution of the crowdsourcing to address the need of solving complex and knowledge intensive tasks, and as means to improve the quality of crowdsourced input. We therefore propose augmenting learnersourcing tasks with expertsourcing tasks within automated knowledge acquisition workflows to not only scale the process of semantic annotation and interlinking, but also ensure high quality annotation and reliability.

We therefore leverage learners, with varied skills and background knowledge, to scale this process of semantic annotation and interlinking. We present the design and implementation of a composite learnersourcing workflow which is composed of two levels. At the first level, the workflow is divided into three-stages: Lookup, Search and Verify-Annotate. Each stage in-turn consists of and interleaved sub-workflow of learner and expert tasks. We introduce expertsourcing tasks, tightly integrated within the learnersourcing tasks, for validating learner contributions and ensuring high annotation quality and reliability. Quantitative measures of ensuring annotation quality are woven into the very fabric of learnersourcing. Our workflow design in driven by semantics based task specifications and learner profiles based on which we devise a mechanism for adaptively selecting the adequate number of learner contributions in order to find the right balance of learners and experts for maximizing learner input and minimizing expert time and contributions. We also introduce the concept of a dynamic task lifecycle manager, that manages the tasks' transition through the learnersourcing and expertsourcing stages.

We demonstrate the effectiveness of our workflow through a prototype implementation of the system tested with upto 30 learners. We present the results of this study and show that we can generate high quality annotations by engaging learners, while reducing the need and contributions from experts. Our workflow design claims to reduce cognitive overload by decomposing complex task into atomic tasks with manageable complexity. Different learners with different skill level can engage by contributing a little even if they cannot manage the whole task on their own. We perform analysis on learners' performance and also reflect on the learnability aspects derived from the process.

7.2 Background and Related Work

We first present some details about the context for which the study is designed. We then present related work in the areas of formalized knowledge modeling for the Islamic domain and towards the intersection of semantic web and human computation methods which forms the basis of this research.

7.2.1 Research Context

In order to understand the concept of inter-contextual and thematic interlinking and to appreciate the need for it, it is extremely important to understand the principles of Qur'anic understanding and learning that are adopted at Islamic institutes across the world. These are summarized here as derived from [154] and illustrated in Figure 7.1.

Linguistic Understanding: Understanding and explanation of the linquistic meanings (of the Arabic Text).

Explanation of the Verse in the light of other verses: This is the foremost necessary condition for establishing the most correct understanding of a given verse - called the Tafseer (Exegesis) of Qur'an by Qur'an.

Understanding the verse and its context using the Hadith (Prophetic Narrations): This is the second most important principle of Qur'anic understanding. This involves understanding the historical context of the revelation of the verses and is often derived from the narrations of the Prophet (which are recorded in the books of Hadith). Also, the explanation of various terms and methodologies have been explained only in the Hadith and it is not possible for a student of knowledge to understand them without understanding the related Hadith.

The foremost contribution of this chapter is to use the learnersourcing methodology, a specialized form of crowdsourcing, using the custom developed web application to gather annotations, classifications and contextual relationships between the Qur'anic verses, themes and the Hadith. The scope of this study is limited to *Verse to Hadith Links* (Inter-contextual links between the Qur'anic verses and Hadith), shown by the link between Verse B to Hadith Z in the Figure 7.1. Capturing these links, primarily from scholarly interpretations, is by far the most knowledge intensive and complex task, since the process may not be automated. A key challenge in this task is that the relations must not be based on mere understanding of students or the teachers, rather attributed to a well known authority mentioned in the books of scholarly works.

7.2.2 Formalized Knowledge Modeling for the Islamic Domain

The challenge of capturing Qur'anic knowledge formally has been captured also by Atwell and collegues [18]. The learners of Qur'anic knowledge face considerable problems in retrieving and synthesizing the most relevant knowledge from a plethora of disparate knowledge sources, given there are neither any formal means of thematically and contextually interlinking them, nor any specialized tools or applications that are designed with the principles outlined earlier. The most commonly available web-based applications for the Qur'an and Hadith provide browsing of the text and available translations, listening to its recitation, word to word meanings, arabic word roots and morphology. Some detailed works of exegesis are also available. However, the use of context specific knowledge discovery tools is limited, given that there is very little or no formalized thematic classification and linking.

Thematic annotation of the Qur'anic verses and Hadith is a non-trivial task. The best known annotations of the Qur'an only focus on its morphological aspects [59]. A recent open linked dataset called 'Semantic Quran' has been published [168], with more than 48 known translations obtained from Tanzil[2]. Pure computational means such as text mining and machine learning techniques fail to provide such classifications, because the nature of the classification and the links is highly sensitive and knowledge intensive. Therefore, human contribution is considered an imperative. We believe our work is the first of its kind that proposes formal

[2]http://tanzil.net/

modeling and acquisition of thematic links between the Qur'an and the Hadith.

7.2.3 Interplay of Semantic Web and Human Computation

One of the major challenges hindering successful application of knowledge driven and ontology-based approaches to data organization and integration in specialized domains is the so-called *'knowledge acquisition bottleneck'*[162]- that is, the large amount of time and money needed to develop and maintain the formal knowledge models and ontologies. This also includes ontology population and semantic annotation using well established vocabularies. To overcome this so-called knowledge acquisition bottleneck, semantic web researchers have recognized that the realization of the semantic and linked data technologies will require not only computation but also significant human contribution [186],[182]. Humans are simply considered indispensable [54] for the semantic web to realize its full potential. There has been growing interest to use crowdsourcing and human computation methods to support the semantic web and linked open data research by providing means to efficiently create research relevant data and potentially solve the bottleneck of knowledge experts and annotators needed for the large-scale deployment of semantic web and linked data technologies. Studies such as Inel et al. [100], Noy et al.[148], Sarsua et al.[164] and others have attempted to solve the bottlenecks of human experts and the needed human contribution in the semantic web development (ontology engineering) processes such as semantic annotation and data interlinking. Some early efforts that led to the evolution of this approach include Ontogame [184] and inPho [147]. Two major genres of research may be seen emerging in the last few years, in an attempt to bring human computation methods to the semantic web: 1) Mechanized Labour and 2) Games with a Purpose for the Semantic Web. Several recent research prototypes have attempted to use micro-task crowdsourcing for solving semantic web tasks e.g. ontology engineering [148], [164] and linked data management [180]. Recent work by Hanika, Wohlgenannt and collegues [5, 81] have attempted to provide tool support for integrating crowdsourcing into ontology engineering processes by providing a plugin for the popular ontology development tool Protégé. Other approaches such as [183] and [173] adopted the Lui von Ahn's "games with a purpose" [202] paradigm for creating the next generation of the semantic web. The idea is to tap on the wisdom of the crowds by providing motivation in terms of fun and intellectual challenge.

Research also suggests the use of ontologies to improve the human-computation process [125], which provides the motivation for proposing a semantics driven model of human computation for this research.

7.2.4 Learnersourcing and Crowdsourcing Workflows

The success of crowdsourcing [102] workflows in general and learnersourcing workflows [209] in particular has motivated this research, to harness the power of human computation by embedding it within learning scenarios in the Islamic domain. Researchers recognize that much work in the real world is not amenable to crowdsourcing because of the difficulty in decomposing tasks into small, independent units [77]. We also recognize this problem in our research and propose decomposition of complex tasks into manageable units, leveraging on skills of different types

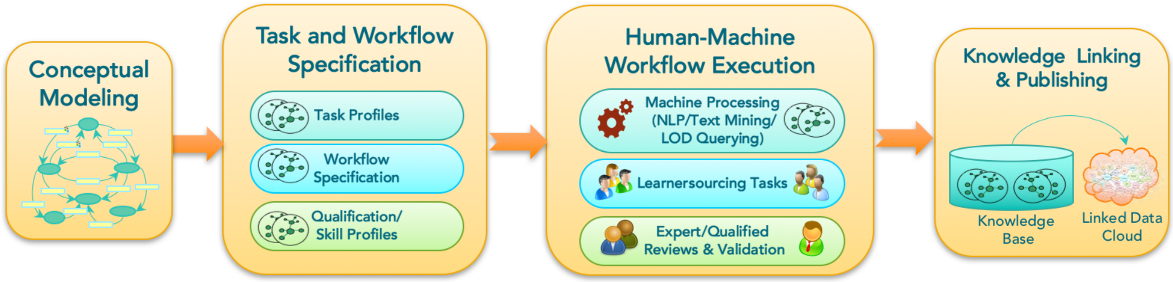


Figure 7.2: Overview of the Learnersourcing Methodology

of learners.

7.3 Research Approach: Semantics Driven Learnersourcing Methodology

7.3.1 Approach for Harnessing Students' Learning

When the students are engaged in Qur'anic studies, they are asked to produce an analysis or a synthesis of the learning material for a given theme or a subject. In doing this, they naturally follow the learning or knowledge discovery path as outlined in the Figure 7.1. We aim to channel this natural learning effort into tasks that capture this knowledge and store it into structured and formalized knowledge models. We also make use of Natural Language Processing (NLP), Information Retrieval (IR) and Text Mining methods, and the available data sources to facilitate this process of knowledge discovery.

We aim to harness not only the learners on the online and distance learning portals, but also those in more traditional classroom environments and introduce the notion of electronic assessments and assignments, if such a system is not already in place. In the case of online learning environments, where some form of e-learning assessment methodology is inplace, we plan to augment it using the methodology described in this chapter. For the initial study presented in this chapter, we implement a prototype system of our own, called Qur'anic Informatics [3], which will be used to introduce supplementary assessments and assignments into the learning methodology of both online and traditional classrooms.

7.3.2 Overview of the Learnersourcing Methodology

The high-level view of the learnersourcing methodology we developed is shown in Figure 7.2. Conceptual knowledge models (ontologies) are used to drive the tasks and workflow specifications. A typical learnersourcing workflow in our system refers to a series of machine and human tasks interleaved together. We introduce the notion of *task profiles* and *learner profiles*, which allows for introducing different tasks in different learning scenarios. It also facilitates leveraging different learner types and levels, according to knowledge background and skills. Learners are given *qualification profiles* for this purpose. Another key feature of the workflow is the *response aggregation, annotation agreement* and *validation.* The tasks which do not meet the desired agreement thresholds are passed on to advanced learners or experts (teachers or scholars), who validate and review the tasks. We also

[3] www.alasmaalhusna.islamicinformatics.org

introduce interleaved machine tasks based on NLP and textmining methods where appropriate. This has been done to enhance the user contribution in order to provide some seed input. We use two available datasets for this purpose[4]. We utilize an available thematic hierarchy [5] for providing some seed themes to the learners and for selecting themes pertaining to a given verse as illustrated in Figure 7.3. The last stage of the method includes publishing the validated annotations as linked and reusable knowledge, for which we aim to use established semantic web and linked data [6] standards.

7.4 Dynamic Learnersourcing Workflow Design for Thematic Linking and Annotation of Hadith

We design our workflow for this study in the light of a case study with the focus on thematic and inter-contextual linking of the Qur'an and the Hadith. Before describing the workflow design in details, we outline some factors that make the Hadith interlinking challenging, and describe some formative design activities that guided the final workflow and task design.

7.4.1 Challenges in Workflow and Task Design

Interlinking the Qur'anic verses and the Hadith is a non-trivial task. Basharat et, al. [30] highlight some of these challenges and propose a linked data vocabulary for

[4]www.quranontology.com, corpus.quran.com
[5]http://quranytopics.appspot.com
[6]http://linkeddata.org

modeling these links. We summarize some factors that make this process challenging. There are number of well known Hadith repositories available, which provide the provision of browsing and searching the Hadith collections such as sunnah.com, dorar.net being the most prominent ones. Most of the classical knowledge sources do not use a standardized numbering scheme for the Hadith. Therefore, it is difficult to link them to the modern Hadith repositories. This is unlike the Qur'anic verses which have a standardized numbering scheme. There are multiple sources of the Hadith, which may have different levels of authenticity which is a matter of discussion beyond the scope of this research. Despite the fact that most Hadith collections have now been classified into authentic categories, the mapping of this classification to the sources that cite them is only possible if the Hadith are extracted and linked in a formalized manner. In addition, to add to the challenge, the Hadith are of varying length, and oftentimes the commentator or the *tafsir* scholar will only quote a part of the Hadith or make a passing reference to it, making it extremely difficult to trace the original Hadith being cited. Moreover, several Hadith may have common portions of narrations, therefore it makes it all the more challenging to identify, which exact Hadith is being quoted or referred to. To overcome some of these challenges, we therefore propose automated knowledge acquisition workflows for acquiring these annotations and links.

7.4.2 Formative Design Activities

The main high-level task identified to be learnersourced as part of this case study was identifying a Hadith or a portion of it referenced by a scholar in a book of

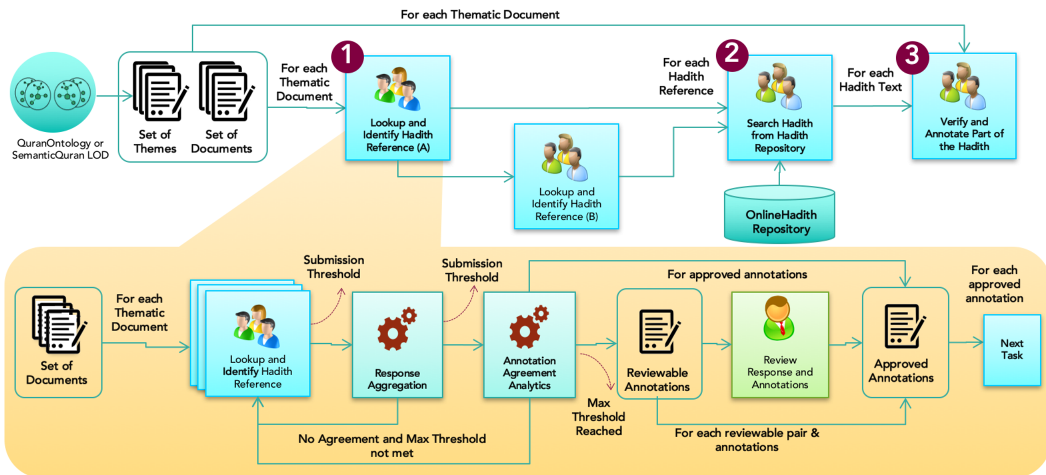


Figure 7.3: Workflow Design

exegesis or scholarly commentary, linking it to a standard Hadith repository (in this case sunnah.com), and annotating the part of the Hadith which is referenced by the scholar, if the entire Hadith has not been referenced. This can prove to be a time consuming task. The typical steps involved in identifying, searching and retrieving/selecting a Hadith by a researcher are as follows: If the Hadith number is available and cited in the source, along with the book number, the Hadith is looked up and verified. If the Hadith number is not available, a search is performed using some of the words from the Hadith. Several Hadith may be returned as a result of a single search. The researcher now tries to match the Hadith with the original text and retrieves the link. The researcher notes the link and the reference of the correct Hadith. In case there are multiple Hadith, a note is made for the different sources.

We designed a first prototype of the task was designed as a single task with several detailed steps. It was shared with upto 10 participants to obtain feedback for usability and ease. The participants reported that the task comprised of too many steps, one of which proves quite challenging and time consuming. This involves searching the Hadith and locating the correct one. One participant suggested using alternate method to locate the Hadith i.e. using the Hadith number. While this method may not always be applicable, we decided that for the first pilot we would limit our study to capture only numbered references so as to simplify the task to some degree and also be able to measure task performance.

7.4.3 Composite Workflow Design Decisions

Based on the feedback from the formative design phase, we felt a strong need to decompose the task into constituent units that impose less cognitive overload and incur less time and effort for the completion. We conceived and designed a composite task workflow. The level of granularity of its constituent sub-tasks and workflows was a design decision that involved several formative stages. A design goal in designing these assignment tasks was to ensure that the input or the response soliciting method aligns with the learners' natural path of learning, which they are familiar with and are inline with their natural enquiry. We designed and executed the workflow in two distinct phases. The results and the design feedback from the Phase 1 was used to iterate and make design modifications for Phase 2. For all tasks, the learners are guided via step by step instructions and tutorials are provided for any task which may require additional steps such as searching the

Hadith from the online Hadith repository or locating the Hadith reference from the document that is linked.

7.4.4 Three-Staged Composite Workflow Design - Phase 1

The main workflow is shown in 7.3. The workflow is designed as a composite workflow consisting of three sub-stages. Each stage in-turn consists of learner, machine and expert tasks interleaved together. For each stage, we describe how the learner tasks are generated, the task interface design, the decision metrics, profile parameters and when and how expert reviews are sought.

Stage 1 - Lookup and Identify Hadith Reference

Task Generation: This task takes as input a set of documents which may or may not contain Hadith references relevant to the theme of the document. For this case study we took 97 themes and generated the task instances using these as seed input. Similar themes have been used in [32].

Task Design: The task is divided into three steps. The first step requires the learner to open the linked document, with relevance to a theme and lookup any Hadith reference. The first part of the task interface design is a simple Yes or No question, asking the user to identify if a Hadith reference is found in the given document. Only if the user answers Yes, the second part of the task is dynamically generated asking the user to identify the source book of the reference. A semblance of this task is shown in Figure 7.4. The user is required to enter a numeric reference to the Hadith and is also required to specify if this is the only

Hadith reference present in the document. It is worth mentioning that a huge body of knowledge and classical literature, the texts do not contain numbered Hadith references, or they are in a format which is not easily extractable via machine learning approaches. The workflow is meant to extend to include such instances in the future.

The nature of the references contained in these documents could be one of the following:(1) A numbered reference from Sahih Al-Bukhari or Sahih Al-Muslim, two of the most prominent,well-acknowledged and widely accepted Hadith collections of the highest authority. (2) References that are attributed as 'Agreed Upon by both Bukhari and Muslim' (3) References from other books. For the first phase of the case study we focused on the first two types of references. This was done for these primary reasons. First, the references are of higher authority and are more relevant for capturing. Second, the references have a more standardized numbering across the collection and are easily locatable. Last, but not the least, more number of learners are familiar with these books as compared to the other books. For the first two reference types, user selects from closed-ended options and provides the numeric reference in the next step. For the third reference type, response type is open-ended. The user is required to provide the book name and number as given in the document.

Task Life Cycle and Decision Metrics: Each task is given a minimum and maximum *response threshold*,T_{MIN} and T_{MAX} respectively. T_{MIN} indicates the number of unique submissions this task must receive before it undergoes any evaluation, analytics or expert review. Once the minimum number of submissions have been received, *agreement analytics* is performed on the submissions. This has two parts: aggregating the responses to obtain *Response Aggrement (RA)* and the *Annotation Agreement (AA)*. Response agreement involves determining if an agreement is reached, based on a probabilistic *Response Agreement Threshold,* T_{RA}. If this threshold is met with the obtained number of responses, then annotation agreement is determined. However, if the response agreement threshold is not met, the minimum response threshold is increased. What this means is that more submissions are sought for this task. This process is repeated until the maximum response threshold, T_{MAX} is reached. Once this happens, and no agreement has been reached, the task is marked for expert review. For annotation agreement analytics, a similar decision process is applied. An *annotation agreement threshold.* T_{AA} is specified. If this is met, the task is considered accepted and complete and the annotation is used to create an instance of the next task in the workflow i.e. Search Hadith. However, if an agreement cannot be reached from amongst the submissions, and the maximum assignment threshold is met, the task is marked as reviewable and it is considered open for expert review. This process is illustrated in Figure 7.6.

Task Profile Parameters: The task profile parameters specify the minimum and the maximum number of submissions required for task completion. It also specifies any specific qualification profile to be possessed by the learners. In addition, it specifies how many expert contributions are sought for this task, if needed. For this task, the minimum submission threshold is set as 3, and the maximum as 7. Only 1 expert review is required if needed.

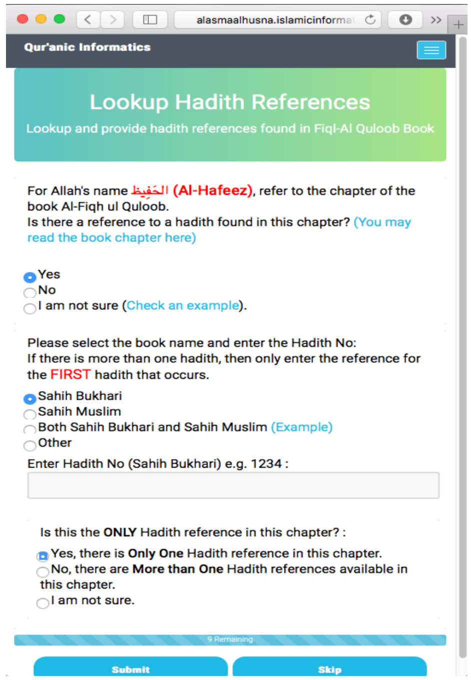


Figure 7.4: A snapshot of the look up hadith task, divided in three dynamic steps

Expert Review: An expert review for this task is required if the maximum assignment threshold is reached and an agreement has not been reached. The experts are invited to use the application using a special invite code and their profile. Both the learners and the experts use the same task interface. However, when an expert logs in, he or she is only presented with the tasks that are marked as reviewable. An expert is able to see the submissions from all the learners and may either select one of these as the correct response or provide one of own.

Stage 2 - Search Hadith from Hadith Repository

Task Generation: At least one instance of this task is generated automatically for each task in stage 1 that reaches completion with an acceptance criteria met through the learnersourcing stage. It is also possible that two instances of this task are generated as a result of 1 task completion with acceptance in the first stage. This case occurs when the reference belongs to the 'Agreed Upon' source i.e. the reference occurs in both the collections of Al-Bukhari and Al-Muslim. Both annotations are then used to create the task instances for this stage.

Task Design: This task obtains as input the agreed upon Hadith reference in the form of the book name and number. It requires the learners to search the Hadith on Sunnah.com and identify the shareable Hadith URL and provide it within the given space. Once they do this, our system crawls the URL to obtain the Hadith and presents it to the learner. They are then required to confirm if the Hadith matches the reference they were required to look for. In an event that the Hadith reference is not found, the learners will respond by selecting the option that they were unable to find the Hadith.

Task (Life Cycle) Decision Metrics and Task Profile Parameters: We employ automated verification for a part of this task. We crawl the URL provided by the learner and retrieve the Hadith number and text from this URL. If the retrieved number matches the original reference number which was provided to the learner to search, the task is accepted automatically. Therefore, the minimum submission threshold is kept as 1. If the submission is not accepted for any reason, then the task's submission threshold is increased. This process repeats until the max submission threshold is reached. In this case, this is kept as 3. If any task remains unaccepted and this maximum threshold is reached, then the task is marked as Reviewable. Once the task is accepted, this leads to the creation of an instance of the task 3, which requires the learners to annotate the portion of the Hadith referenced in the text.

Expert Review: The expert review for this task is required once the task reaches the reviewable stage. As with stage 1, the expert will perform the same task as that of the learner. Only one experts contribution is considered to suffice for this task.

Stage 3 - Verify and Annotate Part of the Hadith

Task Generation: This task is generated automatically for each task in stage 2 that reaches completion with an acceptance criteria met through the learnersourcing stage.

Task Design: This task presents a Hadith to the learner and also a link to the document which is expected to contain the entire Hadith or a portion from it. The task requires the user to select one of the three options: 1) The entire Hadith text is referenced in the chapter, 2) Only part of the Hadith is referenced or 3) The Hadith is not referenced. If the learner selects 2) as the option, a textbox is displayed to the user requiring that the portion of the Hadith (in Arabic) may be copied from the provided text and provided in the given field. The first portion of the task is shown in Figure 7.5.

Task (Life Cycle) Decision Metrics and Profile Parameters: The decision met-

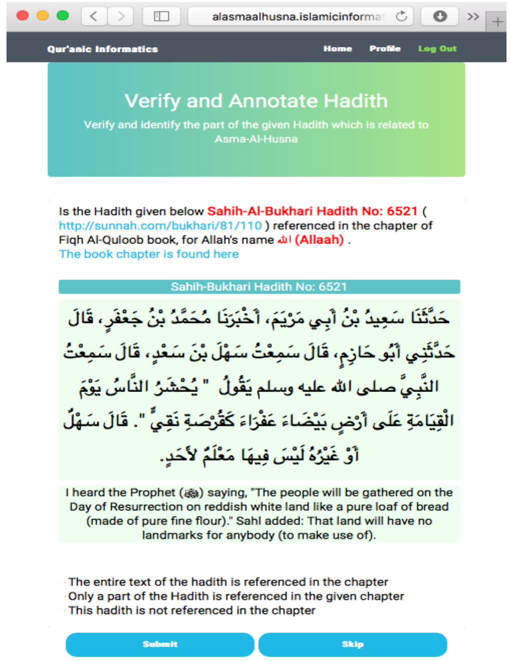


Figure 7.5: A snapshot of the annotate Hadith task, divided in two dynamic steps

rics process for this task is similar to that of task 1. There are 3 submissions required for each task. Once this minimum submission threshold is met, response and annotation agreement analytics are performed. If the task is accepted, then there is no followup except that the state of the task is recorded as completed. If however, the agreement is not reached then more submissions are sought until the maximum submission threshold is reached. If the task is still not accepted then it is marked for review.

Expert Review: The threshold for expert review for this task is still kept to be 1. Since there is a reference document which contains the actual portion of the text, it is not required for more than one expert. However, there may be annotation tasks where it is possible to get different responses. Therefore, there is provision for this threshold to be increased.

7.4.5 Task Design Modification - Phase 2

Following are some of the important design iterations and modifications that were performed in phase 2.

For *stage 1* task, the results of the tasks from phase 1 indicated that the references of the 'other' type never reached an agreement. This was expected as the input type was open ended. In the phase 2, we provided discrete options to the learners, where they could select the book name and then input the numbers. Another important addition to this phase was the creation of followup tasks based on the accepted tasks from phase 1. If the agreed task indicated that there were more references included, then a followup task was created to obtain this second reference. A limitation is that the current workflow does not handle more than two references in the same document, however it is scalable and generic enough to be customized to handle them.

There was no change in the *stage 2* task for phase 2. For the *stage 3*, based on the reviewable tasks from phase 1, it was noted that for some Hadith, there was variation of words in the given chapter vs. those of the Hadith itself. This is not a discrepancy rather a limitation of the question design, since such a scenario may be expected. To cater for this limitation, a fourth option was added to the question's answer options this time. This was to provide better clarity and

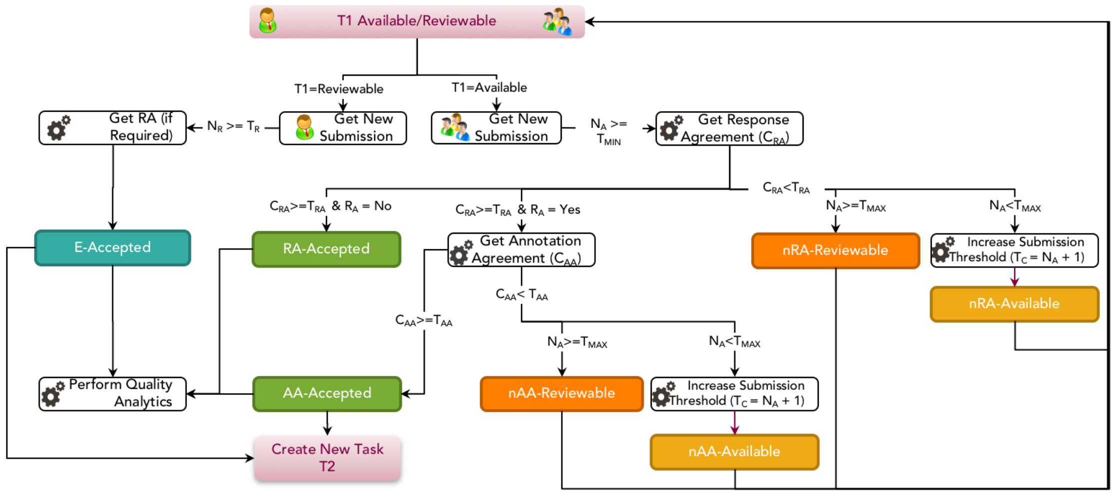


Figure 7.6: Decision Metrics for Task 1

prevent assumptions. Another variation in the question's answer options was made by making the answer option more clear and detailed for the one where the learner selects the whole Hadith being present in the given text. It was noted that some learners in the phase 1 assumed that this excludes the chain of narrators, while others assumed this includes it and selected the second option. While neither is an incorrect assumption, we made the modification to ensure clarity and concreteness in the responses.

7.4.6 Task Life Cycle and Decision Metrics

The Figure 7.7 shows the different stages a task and its submissions may be in. When a task is created its in the *Available* state. This task is available for learners to attempt and submit. Once a learner submits an attempt, a submission is entered in the state *A-Submitted*. Once, the minimum threshold is met for the submissions, the task goes through the analytics stage. It is marked as either *Accepted* or *Reviewable* if the maximum threshold is reached. If there is no *response agreement*, the task is marked as *nRA-Available*, and if the response agreement is reached, while no annotation agreement is reached, then the task is marked *nAA-Available*. On a similar note, if the task reaches a reviewable state without a *response agreement* then, it is marked *nRA-Reviewable* and in the case it doesn't reach annotation agreement, it is marked as *nAA-Reviewable*. A reviewable task will either receive an administrative/controlled review to resolve the disagreement or if deemed suitable it will be sent to the expert. If more than one expert contributions are sought, then similar agreement metrics will be established and task reviewed accordingly. Once the tasks reach agreement, or are reviewed, the original learner submissions will be marked as *Accepted* or *Rejected* based on how they match the majority agreed response or that provided by the expert.

Figure 7.6 shows the process of decision metrics for the Task 1. A unique feature of this decision metric is how the number of available tasks are increased dynamically by applying an adaptive threshold that allows for reaching an agreement with a reasonable margin of error. This model also allows for reducing the number of tasks that will require expert review. If no response agreement or annotation agreement is reached with the minimum response threshold, this threshold is then adaptively changed based on T_{MIN}. If $T_{MIN} <= 5$, T_{RA} and T_{AA} are set as 3. For $T_{MIN} = 6$ or 7, this threshold is increased to 4.

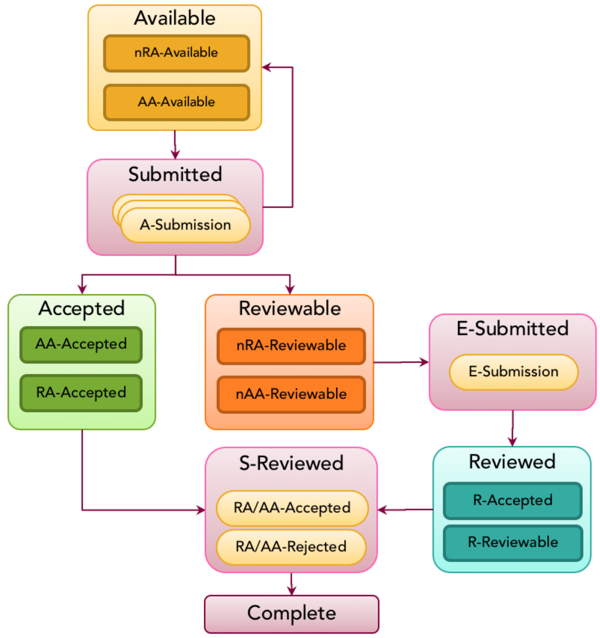


Figure 7.7: Task Lifecycle

7.5 Experimentation and Evaluation

We generated tasks for stage 1 for 97 themes in the Qur'an. We invited learners by open call for participation. Upto 30 learners performed the tasks until they reached completion. For measuring the quality of the proposed system, we performed few substantially different kinds of evaluations: 1) Does the system generate good quality annotations? 2) What is the impact of applying dynamic agreement thresholds and does the system benefit from this method? 3) How are the learner submissions scored in the absence of ground truth? and what is the impact of learners' scores on annotation quality? 4) Does the system facilitate the

Table 7.1: Results for All Tasks

Task State/State	Lookup		Search		Annotate	
	No	%	No	%	No	%
RA-Accepted	45	36.9	89	100	46	50.0
RA-AA-Accepted	60	49.2	NA	-	33	35.9
nRA-Reviewable	1	0.8	0	0	2	2.1
nAA-Reviewable	16	13.1	NA	-	11	11.9
Total	122	100	89	100	92	100

learning process?

7.5.1 Results for Task Acceptance and Annotation Quality

Table 7.1 illustrates the status of the tasks at the end of both phase 1 and phase 2 of the study conducted. The number of tasks fully accepted including those with and without annotations were about 86%. Only 1 task did not reach a response agreement. There were 16 tasks that required a review for annotations. Of these, 8 tasks were in the phase 1, and all these did not have a closed ended answer option as all other tasks. These were actually re-run in the second phase.

For stage 2 tasks, this task had a 100% acceptance rate. For stage 3, the number of tasks fully accepted including those with and without annotations were about 86%. Only 2 tasks did not reach a response agreement. There were 11 tasks that required a review for annotations. All of these tasks were resolved administratively and none needed an expert review. An administrative review lets a super user view all responses from learners and select an appropriate response, when agreement is not reached due to minor differences in annotations.

To evaluate the quality of annotations generated by our system, we compared the leaner generated annotations and responses using administrative reviews in order to evaluate quality measure. It was found out that all the responses generated via inter-learner agreement were indeed correct. In addition, the ones marked reviewable were also of high quality. On most occasions the disagreement was a result of minor bias or assumption. It is worth mentioning that for open-ended annotations, the annotations had to match by 100%. This implies a rather rigid criteria, as there are cases where strict agreement may not always be reached as indicated by the tasks marked as reviewable. In future, we plan to experiment with inter-annotation distance.

7.5.2 Analysis of the Impact of Dynamic Agreement Thresholds

One of the aspects we analyzed was whether the designed learnersourcing workflow benefitted from the adaptive threshold mechanism for selecting the number of learner submissions for each task. Figure 7.8 presents the results for the tasks in stage 1 and 3 where the minimum submission threshold was 3 and the maximum submission threshold was 7. It is obvious that the agreement was easier and quicker to reach for the Lookup task. For both tasks we can conclude that we are able to reduce expert or administrative engagement by atleast 12% for the Look up task and by atleast 28% for verification and annotation task.

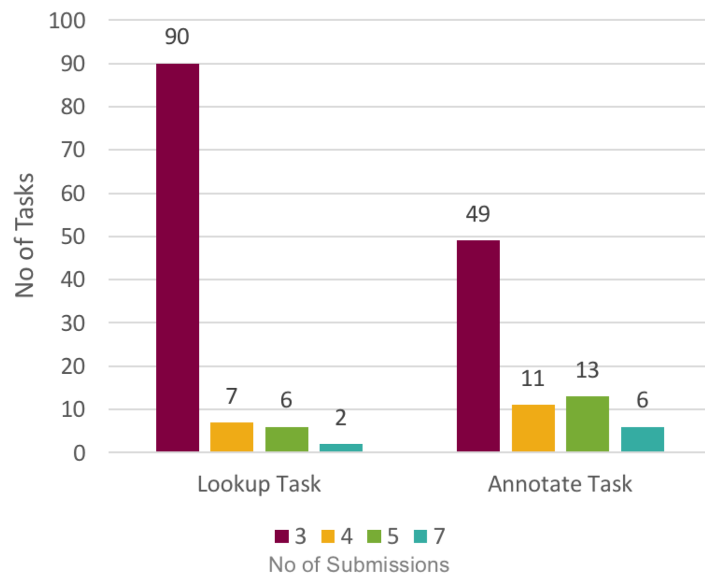


Figure 7.8: Number of tasks that reached agreement for different number of submissions

7.5.3 Scoring and Analysis of Learner Submissions

Once the tasks are Accepted or Reviewed, quality analytics, as shown in Figure 7.6, is performed as part of which learner submissions are marked as accepted or rejected. For this, those tasks that reach an agreement where there is no annotation, the submission that matches the agreed response is marked *RA-Accepted*. If, in addition, a submission contains an annotation, then its classification is added with *AA-Accepted* or *AA-Rejected* depending on if it corresponds with the agreed response. *E-Submitted* reflects that the submission was an administrative or submitted by the expert.

Table 7.2 show the classification for the learner submissions. For stage 1, only

Table 7.2: Learner Submission Classification for all tasks

Submission State	Lookup			Search	Annotate		
	P1	P2	Total	Total	P1	P2	Total
E-Submitted	0	9	9	NA	8	5	13
RA-Accepted	132	0	132	94	118	21	139
RA-Rejected	0	0	0	3	13	0	13
RA-Accepted-AA-Accepted	124	73	197	NA	96	58	154
RA-Accepted-AA-Rejected	2	40	42	NA	32	5	37
RA-Rejected-AA-Accepted	7	5	12	NA	0	11	11
RA-Rejected-AA-Rejected	6	17	23	NA	35	5	40
Re-Runnable	56	0	56	NA	NA	NA	NA
Total	327	144	471	97	302	105	407

5% task submissions are fully rejected. This indicates a high confidence rate is met. The partial rejections account for only 11%. It must be noted that the cases where the annotations are rejected is done fully through the decision metric algorithm. The annotation must fully agree mutually and no credit is given for partial agreements given the concrete and closed ended nature of the task. We also compare the classifications in the phase 1 and phase 2 for this task. A surprising result is a huge number of the cases *RA-Accepted-AA-Rejected*. The lack of cases of *RA-Accepted* in phase 2 is expected, since, this case will only be true when no Hadith reference is found in the text. For phase 2, this is not the case, since all these tasks were followup of the phase 1 and are therefore bound to contain a reference.

For stage 2, only 3 tasks were reviewable. However, these were resolved administratively. The source of Hadith data contained some inconsistencies or difference in the numbering scheme that caused these tasks to be marked as reviewable. For stage 3, 10% task submissions are fully rejected. This indicates a reasonably high confidence rate is met. The partial rejections account for 12%. It must be noted that the cases where the annotations are rejected through the decision metric algorithm, many of those are actually partially correct. It is difficult to apply a concrete acceptance threshold for open ended annotations. The acceptance only reflects complete agreement or with distance 0 between the thresholds. The submissions which correspond to the reviewable tasks are administratively marked accepted. We also compare the classifications in the phase 1 and phase 2. It can be seen that there are no cases of *RA-Rejected* in phase 2. This indicates that the response options provided more clarity to the learners. Also, worth noticeable is that the percentage of cases where both response and agreement are rejected is reduced.

7.5.4 Analysis of Learners' Performance

We also analyzed some of the learners' individual pattern of performance over time, in order to learn aspects of learnability. Figure 7.9 shows two such patterns for the same task. We give numerical classification to learners' submission classification. A score 1 is given if the learners response is accepted, 0.5 if it is partially accepted, 0 if there was no decision, and -1 if it was fully rejected. Learner X's pattern shows that after scoring -1 for 7 consecutive tasks, the learner starts to perform correctly for the task for most subsequent ones. This pattern was also seen for few other learners.

This type of analysis gave us an interesting insight that the system does infact contribute to learnability and the learners are able to learn and improve performance over time. Feedback from some of the learners confirmed this. Some learners reported that they either misunderstood or misinterpreted the question and answered wrongly, only to later realize the error in subsequent questions. Some questioned if they are able to go back and change their responses to the previous questions. However, the system's current design does not provide this provision. Some learners reported that they simply rushed through the responses initially and later realized they answered some questions wrongly. This performance analysis and feedback from the learners also reflects that by imposing inter-learner agreement makes our system robust to accidental and unintentional errors.

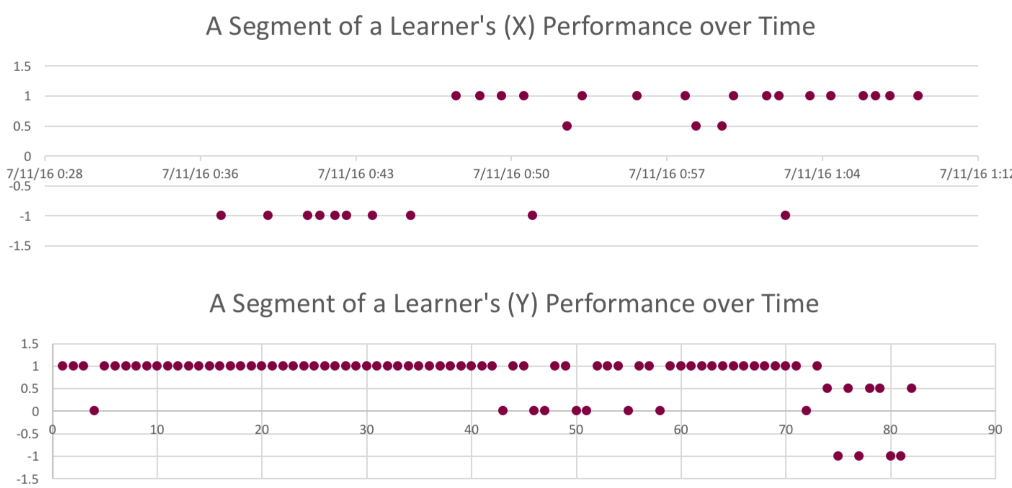


Figure 7.9: Performance patterns for two separate learners for Task 1

The analysis of learner Y's performance provides further interesting insights. This pattern is peculiar because it shows high quality responses for the first 71 responses. However, the last few tasks show partial or complete rejection. Further analysis into this pattern revealed that although the responses were for the same task, the tasks with lower performance were of the third reference type, where the reference was of a complex type. We aim to further formalize these patterns of learner performance, in order to develop adequate measures of learner performance scoring and profiling. These learner and score profiles can play a significant role in further quality control. Also, learners who consistently perform well on a given task may be advanced as super users for the task type. We aim to implement task routing and recommendation based on learners' performance and skill profiles. We believe that this would be crucial to the success of the system when it scales, and other task types and workflows are included.

7.6 Discussion and Conclusions

From the results of the study presented in this chapter, we demonstrated that the adaptive process of increasing learner submissions to achieve inter-learner agreement, for generating thematic annotations, helps scale up the process of annotation without loss of quality. Our results clearly show that the system is robust and the competency of the learners falls within the estimated error margin using which the submission thresholds were generated. The learners performed better than the expected competency in all cases. Depending on the difficulty of the task, the skills of the learners and the availability of the experts these thresholds can be customized appropriately. Additional measures of quality control can be applied by reviewing those submissions by any workers with a consistent pattern of rejected submissions. We plan to infer deeper measures of quality through further administrative reviews.

We were also able to derive important design implications from the iterative design and implementation task workflows. We also learnt that the task design, clarity of instructions, and answer options greatly impacts the outcome. A complex task with many steps is likely to impose cognitive overload for the learners, resulting in reduced engagement and lack of performance. This was handled by decomposing a complex task into independent units whose outputs were later combined into meaningful interconnections and linkages originally sought. The tasks with closed ended and discrete outputs were easier to reach an agreement without administrative or expert reviews and the submissions were easier to classify. However, the open ended tasks required more administrative reviews, in particular, for determining the correctness of the annotations provided by the learners. The annotations are not necessarily wrong, rather, imply a certain bias or an assumption on the part of the learner based on prior knowledge and experience. Increasing the clarity of the question and answer options may help in reducing this bias, so the impact of any prior assumptions involved while attempting the tasks is reduced to a minimal, if not completely eliminated.

In future, we plan to conduct qualitative user studies to assess the enhancement in the learning process that results.

Chapter 8

Semantic Hadith: A Framework for Publishing Linked Hadith Data

لَّقَدْ كَانَ لَكُمْ فِي رَسُولِ ٱللَّهِ أُسْوَةٌ حَسَنَةٌ
لِّمَن كَانَ يَرْجُوا۟ ٱللَّهَ وَٱلْيَوْمَ ٱلْـَٔاخِرَ وَذَكَرَ ٱللَّهَ كَثِيرًا

There has certainly been for you in the Messenger of Allah an excellent pattern for anyone whose hope is in Allah and the Last Day and [who] remembers Allah often.

— *[Al-Quran, Al-Ahzaab,33:21]*

Chapter Overview[1]

This chapter presents ongoing design and development efforts to semantically model and publish the Hadith, which holds a primary position as the next most important knowledge source, after the Qur'an. We present the design of the linked data vocabulary for not only publishing these narrations as linked data, but also delineate upon the mechanism for linking these narrations with the verses of the Qur'an. We establish how the links between the Hadith and the Qur'anic verses may be captured and published using this vocabulary, as derived from the secondary and tertiary sources of knowledge. We present detailed insights into the potential, the design considerations and the use cases of publishing this wealth of knowledge as linked data.

8.1 Introduction

The vast amount of Islamic Creed and legislation derives itself from and is based priamrily on the two most fundamental sources of Islam: namely the *Qur'an* and the *Sunnah* (way of life) of the *Prophet Muhammad*. The later is contained with the vast body of *Hadith* literature [88]. Formally, the *Hadith* is defined as the (recorded) narrations of the sayings and deeds of the Prophet Muhammad.

Our research primarily is motivated to overcome the inherent knowledge acqui-

[1]This chapter is published as:
Basharat, A., Abro, Bushra, I. Budak Arpinar, Rasheed, Khaled, "Semantic Hadith: Leveraging Linked Data Opportunities for Islamic Knowledge", *In the Proceedings of the The Workshop on Linked Data on the Web co-located with the 25th International World Wide Web Conference (WWW 2016),CEUR-WP, Vol-1593, p14.*

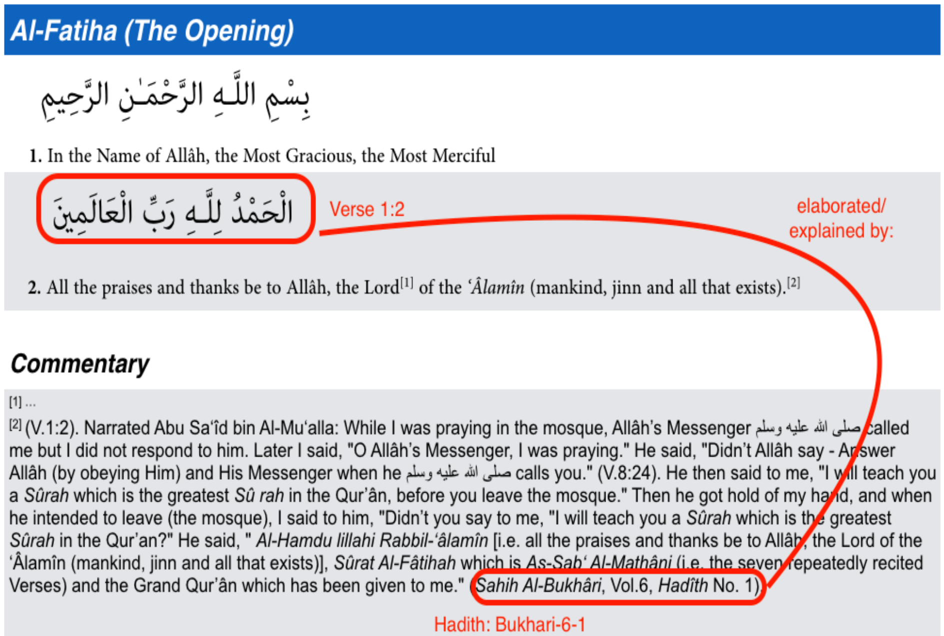


Figure 8.1: A snapshot of a typical Qur'anic Commentary

sition bottleneck in creating semantic content in semantic applications. We have established how this is particularly true for knowledge intensive domains such as the the domain of Islamic Knowledge, which has failed to cache upon the promised potential of the semantic web and the linked data technology; standardized web-scale integration of the available knowledge resources is currently not facilitated at a large scale [29].

8.1.1 Background Context and Motivation

Importance of Hadith

To understand the important of Hadith, the principles of Qur'anic understanding and the science of *tafseer* or exegesis must be considered. The verses in the Qur'an cannot be understood in isolation. The Hadith are used to illustrate the Historical context, the reasons for revelation and elaboration of essential concepts that may not be directly evident. This important principle has been adopted by scholars across centuries to write scholarly commentaries and explanations. Infact, it is a necessary condition to produce an accurate *tafseer* of the Qur'an as explained in detail by Philips [154].

To explain this principle, as an example, consider the Figure 8.1, a derived snapshot taken from QuranComplex[2], the official manuscript, with a translation and a commentary, provided by the Kingdom of Saudi Arabia. The snapshot shows two verses from the first chapter of the Qur'an. The translation is annotated with a commentary (given in the footnotes in this case) in order to provide additional details where important. It is worth noticing that most authentic and reliable commentaries would draw knowledge from the sources of Hadith. In the case of this snapshot, the verse 2 contains an annotation which provides an elaboration based on an authentic Hadith, from one of the many collections of Hadith, called *Sahih Bukhari*, which is known to be the most authentic and reliable Hadith collection.

[2]http://qurancomplex.gov.sa/Quran/Targama/Targama.asp

Motivation: Potential for Knowledge Formalization and Linking

There are hundreds and thousands of Qur'anic commentaries produced over the last few centuries, in various languages that draw upon and rely heavily on the Hadith sources to provide an iterpretation of the Qur'anic verses. Given this fact, the potential for knowledge formalization and linking is not only evident, rather it cannot be overemphasized. Formally modeling this wealth of knowledge and the links would enable new ways of research and knowledge discovery and synthesis - the very motivation for this research. However, realizing this vision to span across the plethora of Islamic resources is a mammoth task. We present some key challenges presented.

Challenges in Interlinking Islamic Knowledge Sources

There have been some recent efforts to publish Islamic knowledge as linked data on the Linked Open Data (LOD) cloud. The efforts primarily focus on the Qur'an. The two datasets that we consider in our research and attempt to link with our Semantic Hadith research include *SemanticQuran*[3] [169] and *QuranOntology*[4] [78].

However, there are no known publically available sources of data or vocabularies published as linked data for the Hadith. There are number of well known Hadith repositories available, which provide the provision of browsing and searching the hadith collections such as sunnah.com, dorar.net being the most prominent ones.

We review some of the state of the art towards computational approaches applied to Hadith texts in Section 8.6. Here, we would like to emphasize that

[3]http://datahub.io/dataset/semanticquran
[4]http://www.quranontology.com

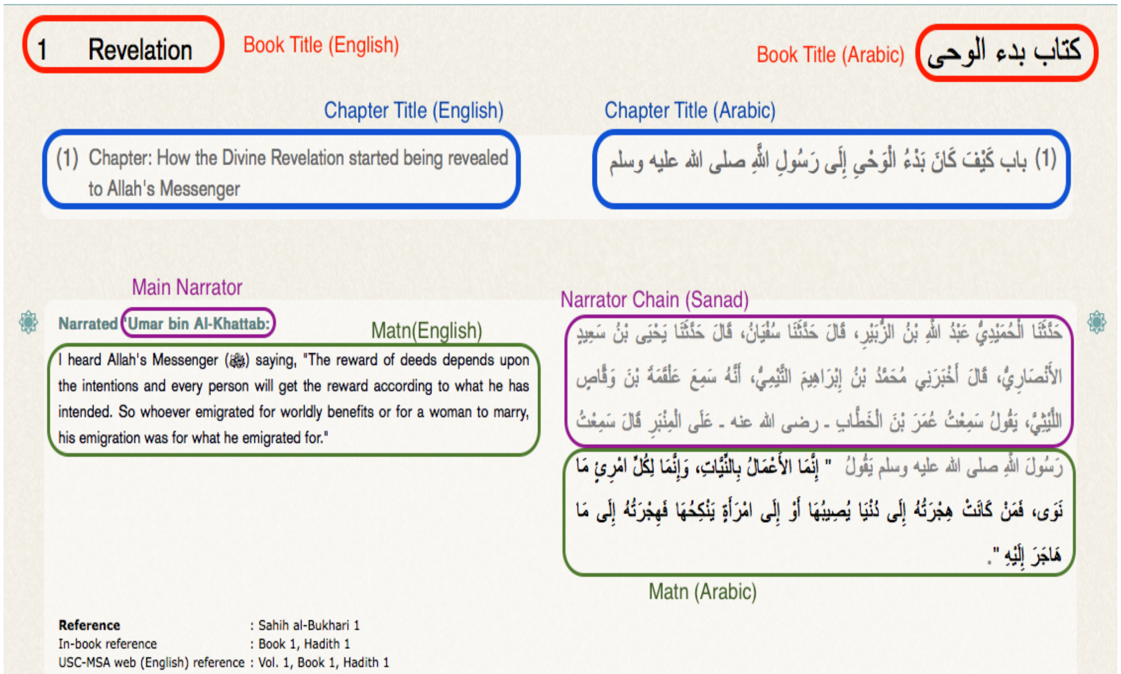


Figure 8.2: A Sample Hadith Snapshot

interlinking the Qur'anic verses and the Hadith is a non-trivial task. We summarize some factors that make this extremely challenging. Most of the classical sources do not use a standardized numbering scheme for the Hadith. This is contrary to the Qur'anic verses which have a standardized numbering scheme. There are multiple sources of the Hadith, which may have different levels of authenticity which is a matter of discussion beyond the scope of this paper. Despite the fact that most Hadith collections have now been classified into authentic categories, the mapping of this classification to the sources that cite them is only possible if the Hadith are extracted and linked in a formalized manner. In addition, to add to the challenge, the Hadith are of varying length, and oftentimes the commentator or the *tafsir* scholar will only quote a part of the Hadith or make a passing reference to it, making it extremely difficult to trace the original Hadith being cited. To add to the challenge, several Hadith may have common portions of narrations, therefore it makes it all the more challenging to identify, which exact Hadith is being quoted or referred to. We believe that a knowledge formalization and linking mechanism, using the linked data standards, is the way forward for solving some or more of these challenges.

8.1.2 Contributions of the Paper

In this paper we make the following contributions:

- We provide the first of its kind linked data model, called *Semantic Hadith* for publishing Hadith as Linked Data and for linking with other key knowledge sources in the Islamic domain, primarily the Qur'an.

- We present a classification of the various levels of links that may potentially be established between the Hadith, the Qur'an and other data sets on the linked data cloud. This classification spans various levels of granularity. We highlight the linking challenges and design issues with each one and present potential modeling solutions.

- We provide a knowledge extraction, linking and publishing framework that may be reused for publishing similar knowledge and linked with the existing linked data cloud. We present our preliminary implementation of this framework.

8.2 Ontology for Semantic Hadith

We first present an illustration of the structure of the Hadith, and then detail upon the design of the ontology for Semantic Hadith.

8.2.1 Hadith Structure

Figure 8.2 shows a sample of a Hadith taken from sunnah.com[5]. A given Hadith has two main parts: the actual narration or the content portion of the Hadith is called *Matan*, and the chain of narrators(reporters) through whom the narration has been transmitted and then recorded is traditionally known as the *Sanad* or simply the *chain of narrators*. The *Sanad* is a chronological chain of narrators, each mentioning the one from whom he heard the Hadith all the way to the prime narrator of the *Matan* followed by the *Matan* itself [157]. The *Sanad* plays the most important role in determining the authenticity of the Hadith, which is the most crucial indicator Scholars resort to when determining whether to accept or reject a Hadith.

8.2.2 Ontology Schema

Figure 8.3 shows the conceptual model for publishing Hadith data on the LOD cloud. Here we summarize the key entities and relations that we chose to include

[5]http://sunnah.com/bukhari/1

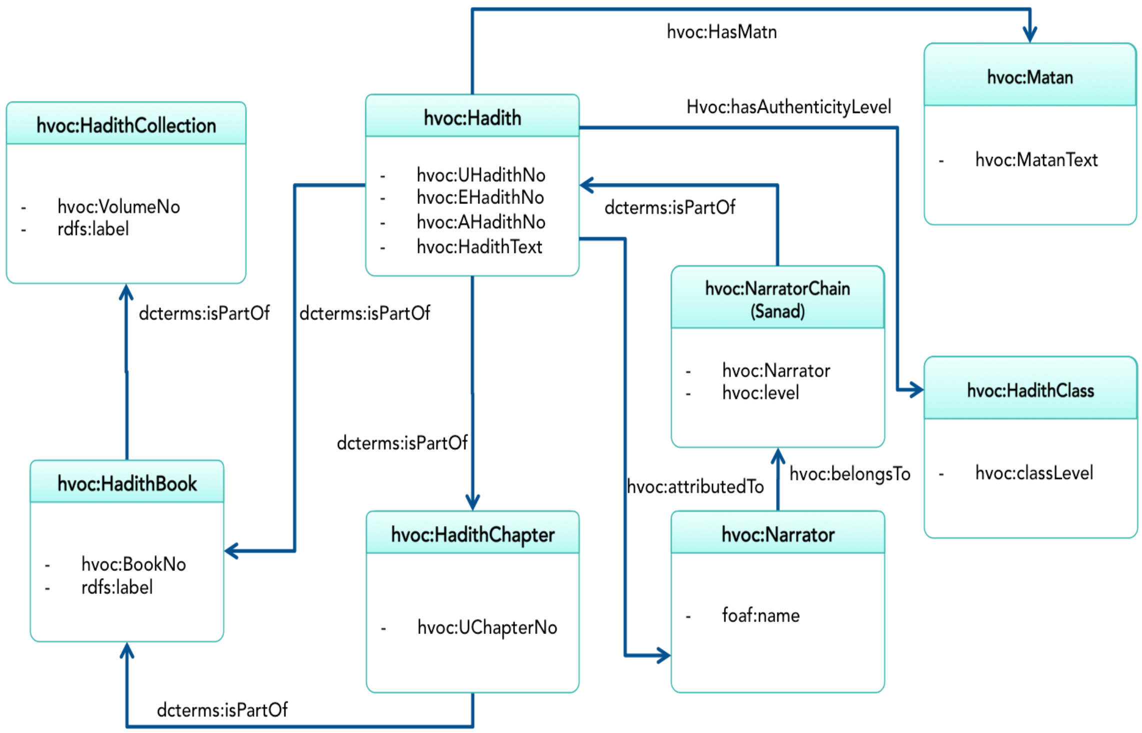


Figure 8.3: High Level Design of Semantic Hadith Ontology

in the conceptual design model of the Semantic Hadith ontology schema.

- *Hadith:* This is the central entity in the domain model. Since there had been no standardized numbering scheme for the Hadith since the beginning, a few alternate numbering schemes may be encountered, therefore the provision to include alternate numberings is made.

- *Matn:* This is primarily a textual entity, which contains the main narration of the Hadith, without the chain or narrators or the Sanad.

- *Narrator:* A Narrator is essentially a Person, with the special role of a narrator of the Hadith. One narrator may have many Hadith attributed to him or her. If a narrator is the root narrator of the Hadith, then a Hadith is usually `attributedTo` him/her. This is shown by the relation between the `Hadith` and `Narrator`. Notice in the Figure 8.2, the english translation does not provide the entire NarratorChain, rather it only provides the name of the narrator to whom the Hadith is attributed to. However, this is not the case for the Arabic (original) version of the Hadith, which usually contains the entire chain of narrators. The chain is often omitted in the books for simplifying the hadith text for the reader and making it more meaningful and relevant. However, the NarratorChain is considered indispensable for determining the validity and authenticity of the Hadith, especially if no other validation source is mentioned.

- *Sanad(NarratorChain):* This is an entity which will contain reference to a Narrator entity, and a level, which will indicate the sequence of the narrator in the chain. Same narrators may appear in many chains.

- *HadithClass:* This indicates the authenticity level of the Hadith. These are detailed in [157].

- *HadithChapter, HadithBook and HadithCollection:* These are entities meant for structural organization of the `Hadith`. A `Hadith` is a `part` of a `Chapter`, which usually contains thematically co-related collections of `Hadith`. `Chapters` are collected in `Books` and `Books` are compiled as `Collection` or Volume.

8.2.3 Vocabulary Design

We choose the `hvoc` prefix for the SemanticHadith vocabulary, as in the domain model. We also ensure reuse of well established linked data vocabularies such as `FOAF`[6] [41], `SKOS`[7] [136], and `DublinCore`[8] [208]. We also provide equivalence relations where applicable. Some of the most relevant equivalence relations are with the `bibo` ontology[9].

8.3 Link Modeling and Design Issues

One of the most important constituents in the design of Semantic Hadith, is the aspect of facilitating the interlinking of knowledge at various levels. We have earlier described the Macro-Structure for Islamic Knowledge in [29]. We distinguish between the nature of links based on the level of granularity at which they are modeled. A *Macro-Level Link* is considered to be one where the source entity is either at the level of a *Verse* in the Qur'an or a *Hadith* in a Hadith Collection. If a link is established for a group of Verses or Hadith, then it will also be considered at the Macro-level. A Micro-level link will be at a sub-verse, sub-Hadith or word or phrase level. For the scope of this paper, we would detail upon only the Macro-level links of the most essential types.

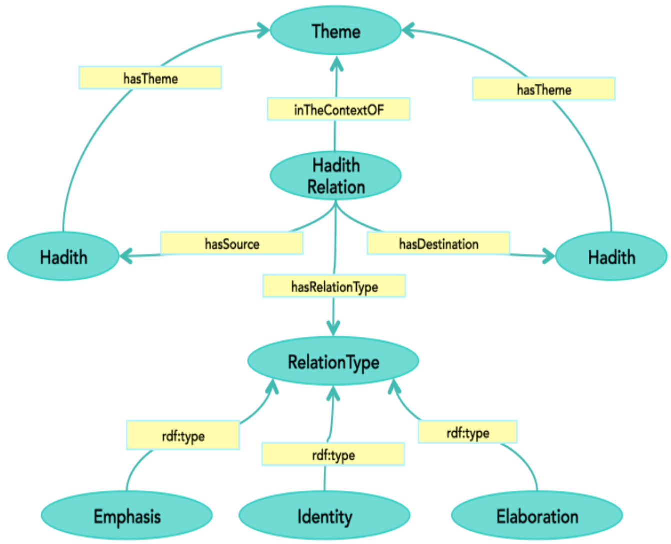


Figure 8.4: Conceptual Design model for Hadith-Hadith Relationship

8.3.1 Hadith-to-Hadith Links

As essential type of links to be established are those links, where by one Hadith is linked to or related to another Hadith. This could be done for Hadith which may be part of the same collection; or it may be between Hadith that are part of different collections. These relations may be of the following primary types: 1) Two Hadith may be considered to be related if they have the same 'sanad'. 2) Two Hadith may be considered to be related if they have the same 'matan'. Note that two Hadiths may occur in the same collection, in two different chapters,

[6]http://xmlns.com/foaf/spec/
[7]https://www.w3.org/2004/02/skos/
[8]http://dublincore.org/
[9]http://bibliontology.com

under different thematic categorizations, however, they may be enumerated or numbered differently. Therefore, by asserting this Hadith as similar/related or identical, we aim to make these links explicit. Oftentimes, the same Hadith may be made part of a different collection and therefore, asserting an identity link would become crucial. This is illustrated in Figure 8.4. To handle the annotations between two Hadith, we define an entity called `HadithRelation`, for which the `source` and `destination` represent the two ends of the relation. The relation would often have a common `Theme`. The `RelationType` indicates whether the two Hadith are similar, indicated by `Identity` as the `RelationType`, or one Hadith may elaborate another indicated by `Elaboration` and so forth. These relation types are not exhaustive and may be iteratively refined.

8.3.2 Linking the Qur'an and Hadith

One of the most significant aspects of linking the Hadith dataset is with the verses in the existing Qur'an datasets. We distinguish two types of relationships that may occur between the Qur'an and the Hadith: 1) There may be Verses, entire of which or part of which may be 'Cited' or quoted in a Hadith. This is the most direct kind of relation that exists between a Hadith and a Verse. 2) The other relations are based on those that can only be derived from Scholarly commentaries. The design and modeling issues for both these types are delineated further.

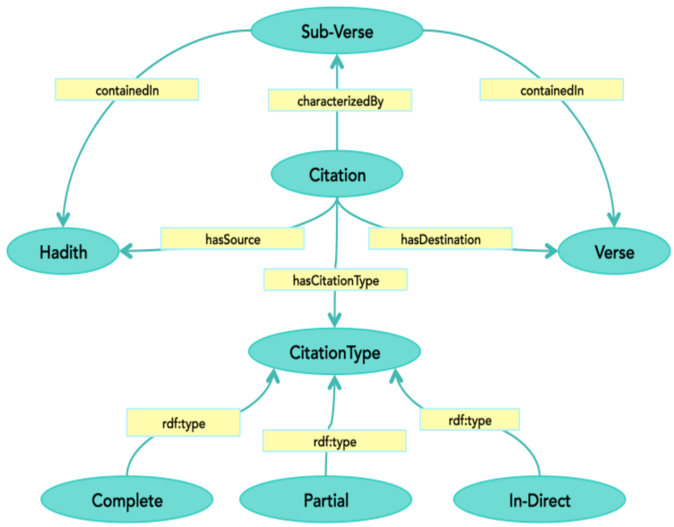


Figure 8.5: Conceptual Design model for Hadith-Verse Relationship

Verse to Hadith Links based on Direct Citations

A direct link between a Hadith and Verse is characterized as one whereby a Hadith contains within its main body a complete verse or a meaningful portion of it. This is modeled in the Figure 8.5. A `Citation` entity is created, which is specific reference to a relation with its `source` as a Hadith and the `destination` as the `Verse`, indicating that its the Hadith that is encapsulating the Verse. It is considered important that we characterize the `CitationType` as either `Complete`, `Partial` or `In-Direct`. A `Complete` Citation will include the entire verse in the body of the Hadith and the Verse will be quoted as such. A `Partial` citation may only contain part of the Verse in the body of the Hadith. To indicate this, the `sub-verse` entity is introduced, which will identify the part of the Verse `citedIn` the Hadith. This is indicated by the relation `characterizedBy`. It is important to note that it is important to annotate and capture the `sub-verse`, since there may be portions of the same verse that may be linked to different Hadith.

Verse to Hadith Links based on Scholarly Commentaries

Another important type of links to be established between the Hadith and the Qur'anic verses are shown in the model as conceptualized in Figure 8.6. This is based on the earlier motivation, provided on the basis of Figure 8.1. In this type of relation, we create an entity `Verse-Hadith-Relation`. In this case, the source is a `Verse` and the destination is a `Hadith`. The reason is that the Hadith will always be used to elaborate or provide the context for the verse in any given commentary or book of exegesis. The `RelationType` may be provided. In this relation type, the most important aspect is establishing the source of the authority of the relation. This is established by the relation `uponAuthorityOf` with a `Scholar` and a relation`establishedIn` with a `Book`. The `Book` is naturally `authoredBy` the `Scholar` to whom the relation is attributed.

8.3.3 Linking Hadith with other Datasources

We aim to provide the provision of linking the Semantic Hadith with other available datasources in the LOD cloud. We present a high level view of the linked cloud model for Islamic knowledge in Figure 8.7. We also mention those datasources, which although are not directly available on the LOD, present potential for linking.

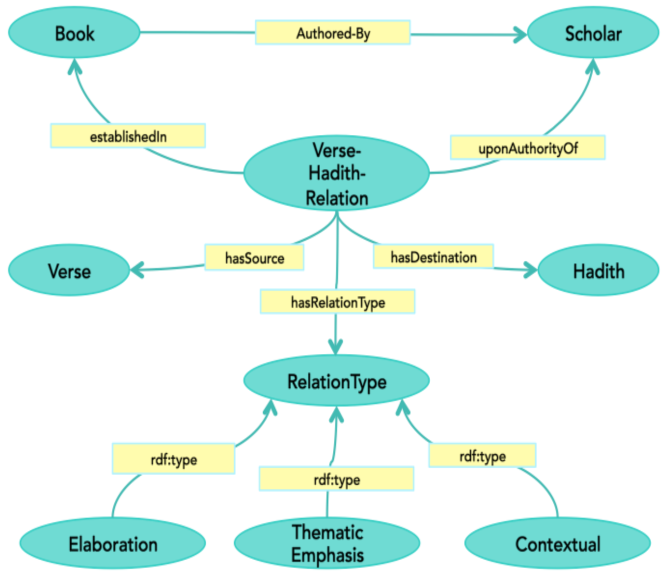


Figure 8.6: Conceptual Design model for Verse to Hadith Relationship based on Scholarly Commentaries

Linking with Existing Datasources in the LOD Cloud

The two available datasources to which the Semantic Hadith is linked to are the `QuranOntology` and `SemanticQuran`. Semantic Quran links itself to DBPedia [10] and Wiktionary [11]. Links would be established between entities in the Hadith to the ones in these two datasets to begin with. For this information extraction would be carried out. There are some important datasources which are not directly part of the linked data cloud but have been made available through QuranOntology and SemanticQuran. These are shown in the Figure 8.7 namely: QuranyTopic-

[10] http://dbpedia.org
[11] http://wiktionary.dbpedia.org/

shttp://quranytopics.appspot.com, QuranCorpus[12], and Tanzil[13].

One of an essential linking aspects would be to thematically map the QuranyTopics to those of HadithTopics.

Linking with other sources

There are other datasources that we plan to link with in the future. Scholars database from a source such as MuslimScholarsDatabase [14] or eNarrator (Hadith Isnad Ontology) [157] [19]. The major limitation is that these sources are not currently available in Linked Data format. However, they present huge potential for linking.

8.4 Linked Data Publishing Framework and Implementation

In order for Semantic Hadith to become a defacto standard and an integral part of the emerging Semantic Web and the LOD cloud for the Islamic Knowledge domain, we also aimed at providing a reusable framework for publishing available Hadith based knowledge sources as linked data. This is shown in Figure 8.8. As elaborated in Section 8.1.1, there are multiple hadith repositories available. Therefore, this reusable framework will benefit multiple hadith publishers to not only expose their data, but also to establish equivalence links with other repositories. This would

[12]corpus.quran.com
[13]tanzil.net
[14]http://muslimscholars.info

be essential towards realizing the vision of linked Islamic knowledge as presented in [29].

8.4.1 Overview of the Framework

The key stages of the framework shown in the Figure 8.8 include: 1) Data Selection, where the data source is selected; 2)Vocabulary Design and Selection, where conceptual and formal knowledge modelling is carried out; 3) Knowledge Extraction, where the process of information and knowledge extraction is carried out; 4) RDF Generation, where the extracted knowledge is converted into the RDF format; 5)Publishing, Linking and Validation is done to make the converted RDF data available via a SPARQL endpoint; and 6) Consumption, is the last stage where the dataset now available as linked data may be consumed into applications.

8.4.2 Implementation Details

We provide some key details of the ongoing implementation process, about the dataset used for publishing as linked data, the knowledge extraction and linking mechanism. We summarize some key results and also highlight some challenges and limitations faced in the implementation process.

Data Sources

As the first Hadith repository to be annotated using the Semantic Hadith Model, we have taken the data of Sunnah.com, which is a structured data repository of some of the most well known and authentic collections of Hadith. The foremost

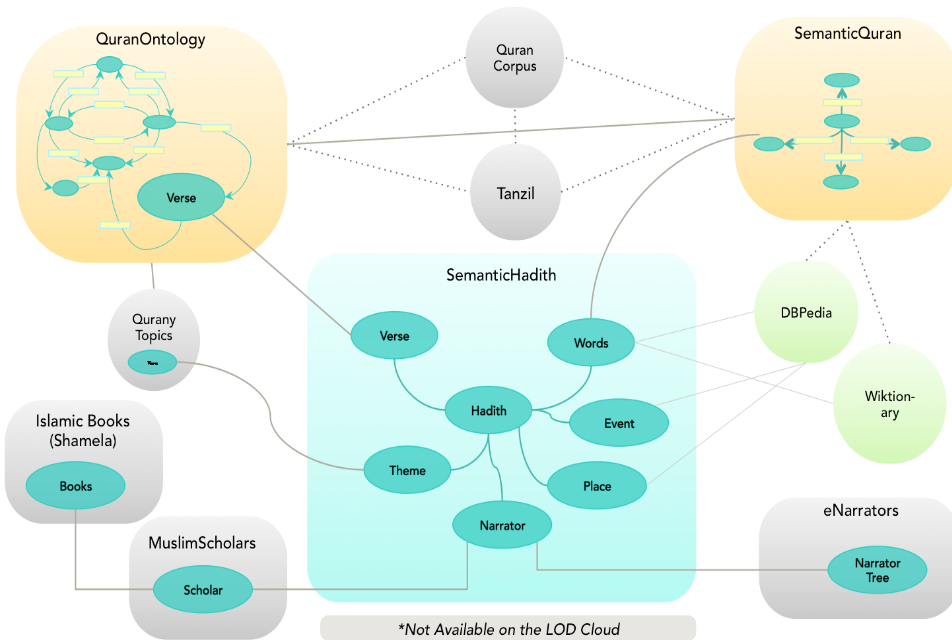


Figure 8.7: A view of the proposed and available Linked Data Cloud for Islamic Knowledge Sources

collections are those of *Sahih Bukhari* and *Sahih Muslim*. Altogether, there are 11 collections in this dataset, with over 25,000 Hadith.

Knowledge Extraction and Linking

For the initial implementation, we focused on extracting some of the key relations explained earlier.

We extracted *Verse-Hadith* Links from QComplex Commentary[15]. This is one of the only datasource through which we were able to extract numbered hadith references, which could be automatically mapped to the hadith collections available with us. An example of such a reference is shown in Figure 8.1. A pattern extraction module was designed to parse the contents of the commentary. The content of the verses, translation and the footnotes were segmented. The mapping between the verses and the corresponding footnotes was easy, given the direct correlation. Pattern matching was then applied to extract the collection name, volume number and the hadith number. This was then mapped to the numbers in our hadith collection. This can be challenging at times, because not all hadith collections use the same type of numbering convention. In such a case, it is non-trivial to map the hadith citation to the corresponding hadith in the repository. Human intervention will be required for validation. We were able to obtain and validate some 300 verse to hadith relations. Since the commentary is not a detailed one, rather comments are only sparingly included as footnotes to the verse translations, it was expected that this number would be small.

[15] http://qurancomplex.gov.sa/Quran/Targama/Targama.asp

We also performed text mining on the arabic text of the hadith data to obtain the *Hadith-Verse* citations, as described in Section 8.3.2. For this, we developed a verse-extraction component, which implements a sub-string matching problem, in order to detect complete or partial verses that may be cited in a given hadith. This is not trivial for several reasons. Different verses span different lengths in the Qur'an. While some may be as long as an entire page's length of a standard book size, others may be as short as one or two words. Therefore, in order to determine, whether the verse is actually being quoted or cited in a hadith requires further validation. Even applying a threshold, relative to the length of the verse, is not an optimal solution. Setting a substantial minimal length was considered, but this may not guarantee a comprehensive coverage. For the first prototype, only 1,325 expert validated links were asserted. In the sunnah.com data, these links may be found as hyperlinks to the verses on the site quran.com.

In addition, similarity computation algorithms were devised to extract *Hadith-Hadith* similarity relations. The 4,973 relations, listed in Table 8.2 are strongly similar Hadith that have at least 60% of text in common. However, the challenge with this approach is that, it cannot be distinguished if the similarity is in the `Sanad` or the `Matan` or both. The more meaningful similarities that are of interest are in the `Matan` of the hadith. In future experiments, we aim to segment the `Sanad` and the `Matan` and extract respective similarity relations. While the similarity threshold for the current approach only took into consideration the common substring, we plan to conduct experiments with more meaningful similarity measures such as Cosine, Jaccard and Pearson correlation coefficient, as done in our work for Qur'anic verses [24].

Results

Based on the dataset and experiments carried out, we summarize some of the dataset and link statistics in the Tables 8.1 and 8.2. Table 8.1 summarizes the statistics for some of the key entities present in the dataset.

Table 8.2 provides the raw count for the candidate relations extracted under the different categories mentioned. It must be noted however, that the relations are not classified according to any of the parameters mentioned in the design. It is also worth mentioning that some of these relations may actually be symmetric.

Table 8.1: Entity Statistics in the Semantic Hadith Dataset(Sunnah.com)

Entity	Count
No of Collections	11
No of Books	311
No of Hadith(Arabic)	25,934
No of Hadith(English)	18,040
No of Chapters	8,968

Table 8.2: Link Statistics in the Semantic Hadith Dataset

Link Type	Count
Hadith-Hadith Relation	4,973
Hadith-Verse Relation(Citations)	1,325
Verse-Hadith Relation(Scholarly)	313

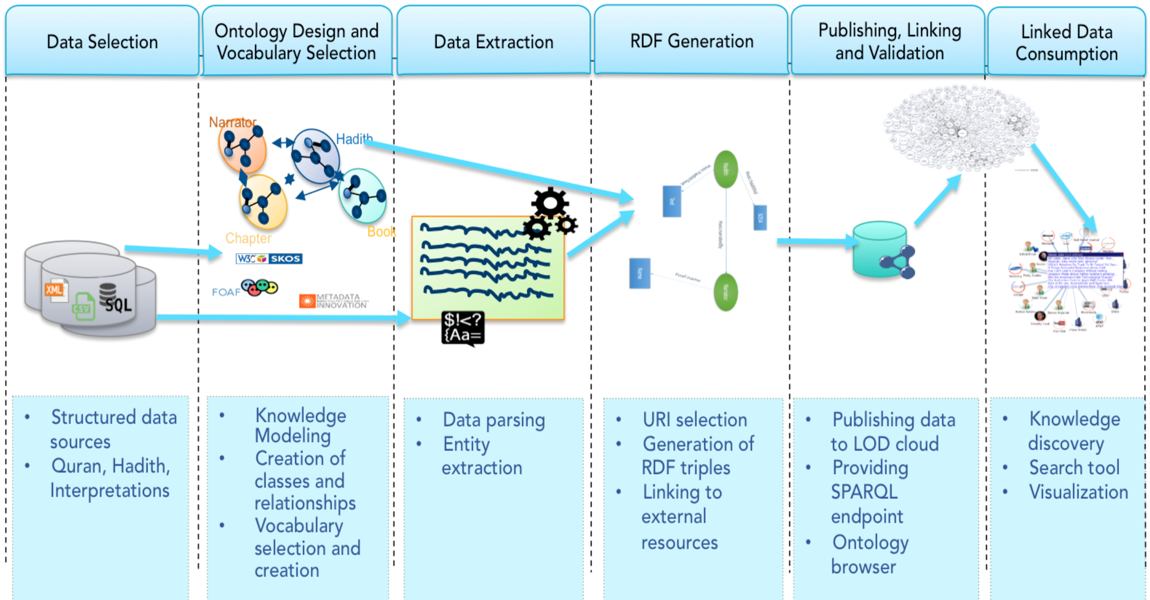


Figure 8.8: Linked Data Generation and Publishing Framework for Semantic Hadith

8.4.3 Existing Limitations and Proposed Solutions

Some of the key limitations we face are with respect to link extraction and validation. There is an obvious lack of structured knowledge sources, with well marked citations. Therefore, the Verse-Hadith links are extremely difficult to be extracted using mere computational means. Human contribution is a must. For this purpose, we intend to pursue a crowdsourcing approach, based on our prior work[28]. We not only intend to use crowdsourcing and human computation methods for the purpose of knowledge acquisition, but also for knowledge validation. Infact, we believe a hybrid human-machine computation methodology to be the only indispensable means of being able to fulfill the vision for linked Islamic knowledge at scale, while ensuring the desired reliability and authenticity.

8.5 Prospective Applications

The most significant benefit of realizing the linked data vision for Islamic knowledge sources will be towards enabling semantics driven distributed knowledge search and retrieval. Most current applications in the Islamic domain only provide limited provision for semantic and conceptual search and retrieval beyond the traditional keyword based searches, upon a single repository. With the Semantic Hadith model, the first of its kind tools will now be possible that would let Qur'an and Hadith repositories to be queried and searched in a federated manner.

The given listing illustrates a federated query between the Semantic Hadith and Semantic Quran datasets. Given that a Verse-Hadith relation exists with the Semantic Hadith dataset, this query retrieves the arabic and english texts for the respective verse.

```
PREFIX rdfs: <http://www.w3.org/2000/01/rdf-schema#>
PREFIX hvoc: <http://www.hadith.islamicinformatics.org/
SemanticHadith#>
PREFIX dcterms: <http://purl.org/dc/terms/>
PREFIX qvoc: <http://mlode.nlp2rdf.org/quranvocab#>

select ?hadith_text ?surahNo ?verseNo
```

```
  ?ayahText ?ayahEng
WHERE {
?verse hvoc:isRelatedTo ?hadith;
 hvoc:verseNo ?verseNo ;
hvoc:surahNo ??surahNo .
?hadith hvoc:hadithId ?hId;
hvoc:hadithText ?hadith_text .

 SERVICE <http://mlode.nlp2rdf.org/sparql> {
?s qvoc:chapterIndex ?surahNo;
 qvoc:verseIndex ?verseNo;
rdfs:label ?ayahText;
rdfs:label ?ayahEng.
FILTER (lang(?ayahEng) ="en" &&
lang(?ayahText) ="ar")
}}
```

This could be taken to another level, by adding another level of federation, and querying the themes of the verse from the QuranOntology.

```
PREFIX rdfs: <http://www.w3.org/2000/01/rdf-schema#>
PREFIX hvoc: <http://www.hadith.quranicinformatics.org/
SemanticHadith#>
PREFIX dcterms: <http://purl.org/dc/terms/>
PREFIX qur: <http://quranontology.com/Resource/>
```

```
select ?hadith_text ?surahNo ?verseNo
?tname
WHERE {
?verse hvoc:isRelatedTo ?hadith;
 hvoc:verseNo ?verseNo ;
hvoc:surahNo ?surahNo .
?hadith hvoc:hadithId ?hId;
hvoc:hadithText ?hadith_text .

 SERVICE <http://quranontology.com/Query> {
?verse qur:DiscussTopic ?t.
?t rdfs:label ?tname.
FILTER(LANGMATCHES(LANG(?tname), "ar"))
}}}
```

This could be further enhanced by automated interlinking with other available datasources on the linked data cloud, as envisioned in Figure 8.7. For instance, once the available hadith are annotated with mentioned events, place or people, they may be linked to the available entities in dbpedia. This would enable richer knowledge discovery and retrieval for a range of applications.

We expect that using this model, more hadith and Qur'anic exegesis repositories, that also rely on and cite heavily the hadith sources, will be published in the linked data format. This will enable the design and development of enhanced learning tools for the Islamic domain, which will provide efficient and personalized access to primary sources of knowledge, ensuring reliability and authenticity. Given that these tools will give better access to meaningfully interlinked knowledge, it will require less effort to find resources and access knowledge beyond books. More content, both classical and contemporary, would become discoverable.

8.6 Related Work

The linked data approach has emerged as the de facto standard for sharing the data on the web.It provides a set of best practices for publishing and connecting structured data on the web [37]. The linked data design issues provide guidelines on how to use standardized web technologies to set data-level links between data from different sources[89]. Increased interest in the LOD has been seen in various sectors e.g. Education [53], [155], Scientific research [16], libraries [137], [95], Government [56], [90], [166], Cultural heritage [63] and many others, however, the religious sector has yet to cache upon the power of the linked open data.

Research in computational informatics applied to the Islamic knowledge has primarily centered around Morphological annotation of the Qur'an [60], [59], Ontology modeling of the Qur'an [7], [21], [65], [211], [212], and Arabic Natural language processing [65]. The LOD take-up in the area of Islamic knowledge has been particularly extremely limited. As mentioned earlier, there have been some recent efforts to publish Islamic knowledge as linked data on the Linked Open

Data (LOD) cloud. The efforts primarily focus on the Qur'an. The two datasets that we consider in our research and attempt to link with our Semantic Hadith research include *SemanticQuran*[16] [169] and *QuranOntology*[17] [78].

Much of the work in the Hadith sciences has focused automating the extraction of the Chain of Narrators. Some works in this regard include [85], [87], [83], [84], [86]. There are also work references with respect to mining the hadith for indexing and classification [9], [145]. Some recent efforts have attempted to model the hadith as semantic ontologies [19] [157]. However, the efforts have focused on annotating the different constituents of the hadith. None of these datasources are available as open source.

Our work is the first of its kind to propose the linked data based model to propose the linking of hadith with the Qur'an. This linked knowledge forms a vital backbone to enable better integration and discovery of knowledge sources.

8.7 Conclusions and Future Work

In this chapter we presented the design and development of our Semantic Hadith framework, which aims to provide the foundation for semantically interlinking the most important Islamic knowledge sources using the linked data standards. We presented the design of the Semantic Hadith Ontology and explained the nature of links with other data sources. The implementation still needs to be matured. The validation of the links and extracted knowledge is a huge challenge we are looking

[16] http://datahub.io/dataset/semanticquran
[17] http://www.quranontology.com

into. We are investigating into crowdsourcing models for knowledge acquisition and validation at scale.

Chapter 9

Conclusion

وَلَقَدۡ يَسَّرۡنَا ٱلۡقُرۡءَانَ لِلذِّكۡرِ فَهَلۡ مِن مُّدَّكِرٖ

> And We have certainly made the Qur'an easy for remembrance, so is there any who will remember?
>
> — *[Al-Quran, Al-Qamar,54:17]*

Chapter Overview

This chapter provides a summary of the work presented in this dissertation and includes a review of the main contributions and conclusive remarks about the research conducted. The scope and directions for future research are also presented.

9.1 Research Summary and Contributions

We believe that this is the first of its kind research that attempts to formally model, interlink and synthesize historical and multi-lingual religious texts using novel human computation methods and workflows. The primary contribution of this research is towards methods that allow for collection and processing large scale semantic annotations from crowd, learners and experts. We also introduce semantics driven interaction techniques to augment annotation and knowledge discovery scenarios for the workers. The method also ensures reliability by introducing 'expert sourcing' workflows tightly integrated within the system. Therefore quantitative measures of ensuring data quality are woven into the very fabric of crowdsourcing and learnersourcing.

The contributions are made possible by uniquely extending and combining the following family of techniques: *crowdsourcing* to aggregate small contributions into meaningful artifacts and *social computing* to motivate participation. We also employ *semantics* driven text analysis, annotation and linking and techniques such as *text-mining* and *linked open data processing* to complement learner input. In addition, we draw inspiration from *learning science* to inform the design of the learnersourcing tasks in order to make the study pedagogically meaningful.

We summarize our key contributions as presented in this dissertation:

- In Chapter 2, we provided insights into understanding the vision, the context, the need and potential for Linked Open Islamic Knowledge; we presented the proposed Macro-structure of linked Islamic knowledge and provides the classification and the nature of links at the knowledge level. We also provided an understanding of the requirements for achieving this vision, and the required foundations by proposing a high-level conceptual framework that would need to be developed to achieve this vision. We also examined the role of Human Computation and how it may be used in overcoming the key challenges associated with resolving the knowledge acquisition bottleneck in order to achieve the LOD vision for Islamic knowledge.

- In Chapter 3, we presented a comprehensive survey of the intersection of semantic web and the human computation paradigm. We adopted a two fold approach towards understanding this intersection. As the primary focus, we analyzed how the semantic web domain has adopted the dimensions of human computation to solve the inherent problems. We present an in-depth analysis of the need for human computation in semantic web tasks such as ontology engineering and linked data management. We provided a 'collective intelligence genome' adapted for the semantic web as means to analyze the threads of composing semantic web applications using human computation methods. As a secondary contribution we also analyzed existing research efforts through which the human computation domain has been better served with the use of semantic technologies. We presented a comprehensive view of the promises and challenges offered by the successful synergy of semantic web and human computation. We also discussed several key outstanding challenges and propose some open research directions.

- In Chapter 4, we evaluated a variety of document similarity measures, using multi-lingual, heterogenous representations of the Qur'an. In order to do this, we developed a similarity computation framework for the verses in the Qur'an. We experimented with four different dataset representations, four similarity measures and three different feature representations and provided comparative insights into the results obtained. Our study proved to be useful for providing seed candidate relations for initiating further crowdsourcing and learnersourcing studies.

- In Chapter 5, we presented the design and development of a semantically enriched task management and workflow generation mechanism applied to the domain of ontology verification, more specifically in the domain of linked data. We devised a task management, workflow specification and automation mechanism that aims to make this process more efficient. This framework forms the foundation of the crowdsourcing module which was utilized in the case studies presented in chapter 6.

- In Chapter 6, we augmented traditional ontology engineering approaches with human contribution through established crowdsourcing and human computation methods. We designed and developed a semantics driven framework, for obtaining thematic annotations from historical and religious knowledge sources. Our framework proposes the use of expertise driven task and crowd profiles for crowdsourcing ontology engineering tasks at varying levels of granularity and knowledge intensiveness. A semantics based task specification model allows for composing automated knowledge acquisition workflows that combine machine computation with crowdsourced annotations at two levels. At the lower level, simple and less knowledge intensive tasks are crowdsourced using the Amazon Mechanical Turk platform. At the higher level for more knowledge intensive tasks, skilled workers and experts are engaged through a custom web application. We demonstrated the effectiveness of this hybrid model for several key knowledge engineering tasks such as thematic disambiguation and thematic annotation, in the Qur'an.

- In Chapter 7, leveraged learners, with varied skills and background knowledge, to scale the process of semantic annotation and interlinking. We presented the design and implementation of a composite learnersourcing workflow which is composed of two levels. We introduced expertsourcing tasks, tightly integrated within the learnersourcing tasks, for validating learner contributions and ensuring high annotation quality and reliability. Our workflow design in driven by semantics based task specifications and learner profiles based on which we devise a mechanism for adaptively selecting the adequate number of learner contributions in order to find the right balance of learners and experts for maximizing learner input and minimizing expert time and contributions. We also introduced the concept of a dynamic task lifecycle manager, that manages the tasks' transition through the learnersourcing and expertsourcing stages. We demonstrated the effectiveness of our workflow through a prototype implementation of the system tested with upto 30 learners. We presented the results of this study and showed that we can generate high quality annotations by engaging learners, while reducing the need and contributions from experts. Different learners with different skill level can engage by contributing a little even if they cannot manage the whole task on their own. We perform analysis on learners' performance and also reflect on the learnability aspects derived from the process.

- In Chapter 8, we provided the first of its kind linked data model, called *Semantic Hadith* for publishing Hadith as Linked Data and for linking with other key knowledge sources in the Islamic domain, primarily the Qur'an. We also presented a classification of the various levels of links that may potentially be established between the Hadith, the Qur'an and other data sets on the linked data cloud. This classification spans various levels of granularity. We highlighted the linking challenges and design issues with each one and present potential modeling solutions. We also provided a knowledge extraction, linking and publishing framework that may be reused for publishing similar knowledge and linked with the existing linked data cloud. We presented our preliminary implementation of this framework.

9.2 Research Significance and Potential Impact

In this dissertation, we proposed a semantics driven framework for combining human and machine computation for the purpose of knowledge engineering. The framework utilizes semantics based tasks, workflow and worker profiles and allows for design and execution of iterative and dynamic workflows based on these profiles. The contextualization that this research aims to tackle, of engineering formalized knowledge models for the purpose of an enhanced knowledge seeking experience in the Islamic knowledge domain, is a high-impact problem, however a non-trivial one. The knowledge at hand is sensitive, and ensuring the credibility and authenticity of knowledge sources is challenging. We demonstrated that a hybrid approach, whereby contributions from crowds and experts, based on skills and knowledge background, combined with automated approaches, can result in creating robust and efficient semantic annotation tasks in specialized domains.

From the outcome of the learnersourcing methodology, we envision a number of novel applications, leveraging the knowledge models that result from the study.

Semantic Knowledge Retrieval and Search: The most significant benefit will be towards enabling semantics driven knowledge search and retrieval.

Enhanced Knowledge Discovery and Learning Tools: This study is one step forward towards enabling enhanced learning tools for the Islamic domain, which will provide efficient and personalized access to primary sources of knowledge, ensuring reliability and authenticity.

9.3 Future Research

We highlight some areas of future research that emerge from the work undertaken as part of this research. This is not meant to be exhaustive list. Several directions have been mentioned in each individual chapter.

- Our initial studies show high quality, robust and favorable results for thematic disambiguation and annotation tasks. We plan to increase the size of the study and experiment with a range of other task designs of varying complexity.

- We also plan to undertake an iterative design approach to revise not only the design of the learner task scenarios, but also the design of the interface and the interaction. We will explore the interplay of better learner methodologies in combination with quizzes and assignments and perhaps group activities for better entailments of thematic classifications.

- While we foresee the investigation into the learning benefits, as an outcome of the learningsourcing efforts, as one of the primary objective towards which future efforts will be directed; a secondary and parallel analysis is planned towards the spiritual and religious effects of the learning achieved, in addition to the pedagogical one. This would give interesting insights into the aspects of 'techno-spirituality'.

- We also envision making the framework generalizable for other knowledge intensive domains.

- There is also need for publishing more data sources as LOD for Islamic knowledge to expand the linked data cloud. On a similar note, applications that consume this knowledge also need to be worked upon.

قُلۡ بِفَضۡلِ ٱللَّهِ وَبِرَحۡمَتِهِۦ فَبِذَٰلِكَ فَلۡيَفۡرَحُواْ هُوَ خَيۡرٞ مِّمَّا يَجۡمَعُونَ

Say, "In the bounty of Allah and in His mercy - in that let them rejoice; it is better than what they accumulate."

— *[Al-Quran, Yunus,10:58]*

Bibliography

[1] URL `https://www.mturk.com/mturk/welcome`.

[2] Noorhan Hassan Abbas. Quran'search for a concept'tool and website. *Unpublished thesis, University of Leeds*, 2009.

[3] M. Acosta, A. Zaveri, E. Simperl, D. Kontokostas, S. Auer, and J. Lehmann. Crowdsourcing linked data quality assessment. *Semantic Web - Iswc 2013, Part Ii*, 8219:260–276, 2013. ISSN 0302-9743.

[4] Maribel Acosta, Elena Simperl, Fabian Flöck, and Barry Norton. A sparql engine for crowdsourcing query processing using microtasks–technical report. *Institute AIFB, KIT, Karlsruhe*, 2012.

[5] Benjamin Adams, Gerhard Wohlgenannt, Marta Sabou, Florian Hanika, et al. Crowd-based ontology engineering with the ucomp protégé plugin. *Semantic Web*, (Preprint):1–20.

[6] Tanja Aitamurto, Aija Leiponen, and Richard Tee. The promise of idea crowdsourcing–benefits, contexts, limitations. *Whitepaper for Nokia Ideas Project*, 2011.

[7] Hend S Al-Khalifa, Maha Al-Yahya, Alia Bahanshal, Iman Al-Odah, and Nawal Al-Helwah. An approach to compare two ontological models for representing quranic words. In *Proceedings of the 12th International Conference on Information Integration and Web-based Applications and Services*, pages 674–678. ACM. ISBN 1450304214.

[8] Mohammed Al Qady and Amr Kandil. Automatic clustering of construction project documents based on textual similarity. *Automation in Construction*, 42:36–49, 2014.

[9] Kawther A Aldhlan and Akram M Zeki. Datamining and islamic knowledge extraction: alhadith as a knowledge resource. In *Information and Communication Technology for the Muslim World (ICT4M), 2010 International Conference on*, pages 21–25. IEEE, 2010.

[10] K. Alexander, R. Cyganiaky, M. Hausenblasz, and J.: Zhaox. *Describing Linked Datasets - On the Design and Usage of VoiD, the "Vocabulary Of Interlinked Datasets"*. in Linked Data on the Web Workshop, Spain, 2009.

[11] O.: Alonso. *Perspectives on Infrastructure for Crowdsourcing*. Workshop on Crowdsourcing for Search and Data Mining, Hong Kong, China, 2011.

[12] UniProt Consortium and. *and others: The universal protein resource (UniProt)*. Nucleic acids research, Oxford Univ Press, 2008.

[13] Lora Aroyo and Chris Welty. Harnessing disagreement in crowdsourcing a

relation extraction gold standard. Technical report, Tech. Rep. RC25371 (WAT1304-058), IBM Research, 2013.

[14] Lora Aroyo and Chris Welty. Measuring crowd truth for medical relation extraction. *2013 AAAI Fall Symposium Series*, 2013.

[15] Lora Aroyo and Chris Welty. Crowd truth: Harnessing disagreement in crowdsourcing a relation extraction gold standard. *WebSci2013. ACM*, 2013.

[16] T. K. Attwood, D. B. Kell, P. McDermott, J. Marsh, S. R. Pettifer, and D. Thorne. Utopia documents: linking scholarly literature with research data. *Bioinformatics*, 26(18):i568–74, 2010. ISSN 1367-4811 (Electronic) 1367-4803 (Linking). doi: 10.1093/bioinformatics/btq383. URL `http://www.ncbi.nlm.nih.gov/pubmed/20823323`. Attwood, T K Kell, D B McDermott, P Marsh, J Pettifer, S R Thorne, D eng BBE0160651/Biotechnology and Biological Sciences Research Council/United Kingdom England Oxford, England Bioinformatics. 2010 Sep 15;26(18):i568-74. doi: 10.1093/bioinformatics/btq383.

[17] Eric Atwell, Nizar Habash, Bill Louw, Bayan Abu Shawar, Tony McEnery, Wajdi Zaghouani, and Mahmoud El-Haj. Understanding the quran: A new grand challenge for computer science and artificial intelligence. *ACM-BCS Visions of Computer Science 2010*, 2010.

[18] Eric Atwell, Claire Brierley, Kais Dukes, Majdi Sawalha, and Abdul-Baquee Sharaf. An artificial intelligence approach to arabic and islamic content

on the internet. *Proceedings of NITS 3rd National Information Technology Symposium*, 2011.

[19] Aqil Azmi and Nawaf Bin Badia. itree-automating the construction of the narration tree of hadiths (prophetic traditions). pages 1–7, 2010.

[20] Chris Baillie, Peter Edwards, and Edoardo Pignotti. Quality reasoning in the semantic web. In *The Semantic Web–ISWC 2012*, pages 383–390. Springer, 2012.

[21] Sumayya Baqai, Amna Basharat, Hira Khalid, Amna Hassan, and Shehneela Zafar. Leveraging semantic web technologies for standardized knowledge modeling and retrieval from the holy qur'an and religious texts. In *Proceedings of the 7th International Conference on Frontiers of Information Technology*, page 42. ACM. ISBN 1605586420.

[22] Linked Movie Data Base(LinkedMDB). URL `http://linkedmdb.org/`.

[23] A. Basharat, I. B. Arpinar, S. Dastgheib, U. Kursuncu, K. Kochut, and E. CrowdLink: Dogdu. Crowdsourcing for large-scale linked data management. pages 227–234. IEEE International Conference on . IEEE, 2014.

[24] A Basharat, D Yasdansepas, and K Rasheed. Comparative study of verse similarity for multi-lingual representations of the qur'an. In *Proc. on the Int. Conference on Artificial Intelligence (ICAI)*, pages 336–343, 2015.

[25] Amna Basharat. Learnersourcing thematic and inter-contextual annotations from islamic texts. In *Proceedings of the 2016 CHI Conference Extended*

Abstracts on Human Factors in Computing Systems, pages 92–97. ACM, 2016.

[26] Amna Basharat, I. Budak Arpinar, Shima Dastgheib, Ugur Kursuncu, Krys Kochut, and Erdogan Dogdu. Crowdlink: Crowdsourcing for large-scale linked data management. pages 227–234. IEEE, 2014. Semantic Computing (ICSC), 2014 IEEE International Conference on.

[27] Amna Basharat, I Budak Arpinar, Shima Dastgheib, Ugur Kursuncu, Krys Kochut, and Erdogan Dogdu. Semantically enriched task and workflow automation in crowdsourcing for linked data management. *International Journal of Semantic Computing*, 8(04):415–439, 2014.

[28] Amna Basharat, I Budak Arpinar, Shima Dastgheib, Ugur Kursuncu, Krys Kochut, and Erdogan Dogdu. Semantically enriched task and workflow automation in crowdsourcing for linked data management. *International Journal of Semantic Computing*, 8(04):415–439, 2014.

[29] Amna Basharat, Khaled Rasheed, and I. Budak Arpinar. Towards linked open islamic knowledge using human computation and crowdsourcing. In *Proceedings of the International Conference on Islamic Applications in Computer Science And Technology*, 2015.

[30] Amna Basharat, Bushra Abro, I Budak Arpinar, and Khaled Rasheed. Semantic hadith: Leveraging linked data opportunities for islamic knowledge. In *Proceedings of the 9th International Workshop on Linked Data on the Web*

(LDOW2016) [held in conjunction with WWW2016], volume 1593, pages 1–10. CEUR-WS.org, 2016.

[31] Amna Basharat, I. Budak Arpinar, and Khaled Rasheed. Leveraging crowdsourcing for the thematic annotation of the qur'an. In *Proceedings of the 25th International Conference Companion on World Wide Web*, pages 13–14. International World Wide Web Conferences Steering Committee, 2016.

[32] Amna Basharat, I Budak Arpinar, and Khaled Rasheed. Leveraging crowdsourcing for the thematic annotation of the qur'an. In *Proceedings of the 25th International Conference Companion on World Wide Web*, pages 13–14. International World Wide Web Conferences Steering Committee, 2016.

[33] D. A. Benson, Mark Cavanaugh, K. Clark, and J. Ostell E. W. Sayers: GenBank Nucleic acids research Oxford Univ Press I. Karsch-Mizrachi, D. J. Lipman. 2012.

[34] Abraham Bernstein. The global brain semantic web - interleaving human-machine knowledge and computation. In *ISWC2012 Workshop on What will the Semantic Web Look Like 10 Years From Now?*, pages 1–6, Boston, MA, 2012.

[35] Abraham Bernstein, Mark Klein, and Thomas W Malone. Programming the global brain. *Communications of the ACM*, 55(5):41–43, 2012. ISSN 0001-0782.

[36] C. Bizer, T. Heath, and T.: Berners-Lee. Linked data - the story so far, int. *Journal on Semantic Web and Information Systems*, 5:3, 2009.

[37] C. Bizer, T. Heath, and T. Berners-Lee. Linked data - the story so far. *International Journal on Semantic Web and Information Systems*, 5(3): 1–22, 2009. ISSN 1552-6283. doi: Doi10.4018/Jswis.2009081901. URL `<GotoISI>://WOS:000270369900001`. 501GS Times Cited:229 Cited References Count:78.

[38] Christian Bizer, Tom Heath, and Tim Berners-Lee. Linked data-the story so far. *Semantic Services, Interoperability and Web Applications: Emerging Concepts*, pages 205–227, 2009.

[39] Kalina Bontcheva, Ian Roberts, Leon Derczynski, and Dominic Rout. The gate crowdsourcing plugin: Crowdsourcing annotated corpora made easy. In *Proc. EACL*, 2014.

[40] Alessandro Bozzon, Marco Brambilla, Stefano Ceri, Matteo Silvestri, and Giuliano Vesci. Choosing the right crowd: expert finding in social networks. In *Proceedings of the 16th International Conference on Extending Database Technology*, pages 637–648. ACM, 2013.

[41] Dan Brickley and Libby Miller. Foaf vocabulary specification 0.98. *Namespace document*, 9, 2012.

[42] Paul Buitelaar, Philipp Cimiano, and Bernardo Magnini. *Ontology learning*

from text: methods, evaluation and applications, volume 123. IOS press, 2005.

[43] L. Carletti, G. Giannachi, and D. McAuley. Digital humanities and crowd-sourcing: An exploration. *Museums and the Web, Portland, Oregon.*, 2013.

[44] Irene Celino, Simone Contessa, Marta Corubolo, Daniele Dell'Aglio, Emanuele Della Valle, Stefano Fumeo, and Thorsten Krüger. Linking smart cities datasets with human computation–the case of UrbanMatch. In *The Semantic Web–ISWC 2012*, pages 34–49. Springer, 2012.

[45] Irene Celino, Simone Contessa, Marta Corubolo, Daniele Dell'Aglio, Emanuele Della Valle, Stefano Fumeo, Thorsten Krüger, and Thorsten Krüger. UrbanMatch-linking and improving smart cities data. In *LDOW*, 2012.

[46] Michelle Cheatham and Pascal Hitzler. Conference v2. 0: An uncertain version of the OAEI conference benchmark. In *The Semantic Web–ISWC 2014*, pages 33–48. Springer, 2014.

[47] Linked Data. - *Connect Distributed Data across the Web*. linkeddata.org/.

[48] J.: Davis. *From Crowdsourcing to Crowdservicing*. IEEE Internet Computing, 2011.

[49] Victor De Boer, Michiel Hildebrand, Lora Aroyo, Pieter De Leenheer, Chris Dijkshoorn, Binyam Tesfa, and Guus Schreiber. Nichesourcing: Harnessing

the power of crowds of experts. In *Knowledge Engineering and Knowledge Management*, pages 16–20. Springer, 2012.

[50] G. Demartini, D. E. Difallah, and P.: ZenCrowd: CudrÃČÂľ-Mauroux. Leveraging probabilistic reasoning and crowdsourcing techniques for large-scale entity linking. *WWW*, 2012, 2012.

[51] Gianluca Demartini, Djellel Eddine Difallah, and Philippe Cudré-Mauroux. ZenCrowd: leveraging probabilistic reasoning and crowdsourcing techniques for large-scale entity linking. In *Proceedings of the 21st international conference on World Wide Web*, pages 469–478. ACM, 2012.

[52] Gianluca Demartini, Djellel Eddine Difallah, and Philippe Cudré-Mauroux. Large-scale linked data integration using probabilistic reasoning and crowdsourcing. *The VLDB Journal*, 22(5):665–687, Oct 2013. ISSN 1066-8888.

[53] S. Dietze, S. Sanchez-Alonso, H. Ebner, H. Q. Yu, D. Giordano, I. Marenzi, and B. P. Nunes. Interlinking educational resources and the web of data a survey of challenges and approaches. *Program-Electronic Library and Information Systems*, 47(1):60–91, 2013. ISSN 0033-0337. doi: Doi10.1108/00330331211296312. URL <GotoISI>://WOS:000316164300004. 106SF Times Cited:1 Cited References Count:91.

[54] Dominic DiFranzo and James Hendler. The semantic web and the next generation of human computation. In *Handbook of Human Computation*, pages 523–530. Springer, 2013. ISBN 1461488052.

[55] L. Ding, J. Shinavier, J. T. Finin, D.: owl:sameAs McGuinness, and Raleigh Linked Data: An Empirical Study, Web Science Conf. Nc (2010).

[56] L. Ding, T. Lebo, J. S. Erickson, D. DiFranzo, G. T. Williams, X. Li, J. Michaelis, A. Graves, J. G. Zheng, Z. Shangguan, J. Flores, D. L. McGuinness, and J. A. Hendler. Twc logd: A portal for linked open government data ecosystems. *Journal of Web Semantics*, 9(3):325–333, 2011. ISSN 1570-8268. doi: Doi10.1016/J.Websem.2011.06.002. URL `<GotoISI>://WOS:000300169800008`. 890UB Times Cited:7 Cited References Count:20.

[57] A. Doan, R. Ramakrishnan, and A.: Halevy. Crowdsourcing systems on the world-wide web. *CACM*, 54:86, 2011.

[58] Anhai Doan, Raghu Ramakrishnan, and Alon Y Halevy. Crowdsourcing systems on the world-wide web. *Communications of the ACM*, 54(4):86–96, 2011. ISSN 0001-0782.

[59] Kais Dukes and Nizar Habash. Morphological annotation of quranic arabic.

[60] Kais Dukes, Eric Atwell, and Abdul-Baquee M Sharaf. Syntactic annotation guidelines for the quranic arabic dependency treebank. In *LREC*.

[61] Anca Dumitrache, Lora Aroyo, Chris Welty, Robert-Jan Sips, and Anthony Levas. Dr. Detective: combining gamication techniques and crowdsourcing to create a gold standard in medical text. In *Proceedings of the 1st International Conference on Crowdsourcing the Semantic Web-Volume 1030*, pages 16–31. CEUR-WS.org, 2013.

[62] Kai Eckert, Mathias Niepert, Christof Niemann, Cameron Buckner, Colin Allen, and Heiner Stuckenschmidt. Crowdsourcing the assembly of concept hierarchies. *Proceedings of the 10th annual joint conference on Digital libraries - JCDL '10*, page 139, 2010. ISSN 9781450300858.

[63] Jeff Edelstein, Lola Galla, Carolyn Li-Madeo, Julia Marden, Alison Rhonemus, and Noreen Y Whysel. Linked open data for cultural heritage. 2013.

[64] E. Estelles-Arolas and F. Gonzalez-Ladron-de Guevara. Towards an integrated crowdsourcing definition. *Journal of Information Science*, 38(2):189–200, Apr 2012. ISSN 0165-5515.

[65] Ali Farghaly and Khaled Shaalan. Arabic natural language processing: Challenges and solutions. *ACM Transactions on Asian Language Information Processing*, 8:1–22, 2009. ISSN 15300226. doi: 10.1145/1644879.1644881.

[66] Dieter Fensel and Mark A Musen. The semantic web: a brain for humankind. *IEEE Intelligent systems*, 16(2):24–25, 2001. ISSN 1094-7167.

[67] Richard S Forsyth and Serge Sharoff. Document dissimilarity within and across languages: A benchmarking study. *Literary and Linguistic Computing*, 29(1):6–22, 2014.

[68] M. Franklin, D. Kossmann, and T.: CrowdDB: Kraska. Answering queries with crowdsourcing. In *Proceedings of SIGMOD*, 2011.

[69] Michael J Franklin, Donald Kossmann, Tim Kraska, Sukriti Ramesh, and Reynold Xin. Crowddb: answering queries with crowdsourcing. In *Proceed-*

ings of the 2011 ACM SIGMOD International Conference on Management of data, pages 61–72. ACM, 2011.

[70] Hanane Froud, Abdelmonaim Lachkar, and Said Alaoui Ouatik. A comparative study of root-based and stem-based approaches for measuring the similarity between arabic words for arabic text mining applications. *arXiv preprint arXiv:1212.3634*, 2012.

[71] David Geiger and Martin Schader. Personalized task recommendation in crowdsourcing information systems–current state of the art. *Decision Support Systems*, 65:3–16, 2014.

[72] Elena L Glassman and Robert C Miller. Leveraging learners for teaching programming and hardware design at scale. In *Proceedings of the 19th ACM Conference on Computer Supported Cooperative Work and Social Computing Companion*, pages 37–40. ACM, 2016.

[73] Elena L Glassman, Aaron Lin, Carrie J Cai, and Robert C Miller. Learnersourcing personalized hints. In *Proceedings of the 19th ACM Conference on Computer-Supported Cooperative Work & Social Computing*, pages 1626–1636. ACM, 2016.

[74] Kwang-Il Goh, Michael E. Cusick, David Valle, Barton Childs, Marc Vidal, and Albert-László Barabási. The human disease network. *Proceedings of the National Academy of Sciences National Acad Sciences*, 104(21):8685–8690, 2007.

[75] Thomas R Gruber. A translation approach to portable ontology specifications. *Knowledge acquisition*, 5(2):199–220, 1993.

[76] Thomas R Gruber. Toward principles for the design of ontologies used for knowledge sharing? *International journal of human-computer studies*, 43 (5):907–928, 1995.

[77] Nathan Hahn, Joseph Chang, Ji Eun Kim, and Aniket Kittur. The knowledge accelerator: Big picture thinking in small pieces. In *Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems*, pages 2258–2270. ACM, 2016.

[78] Aimad Hakkoum and Said Raghay. Ontological approach for semantic modeling and querying the qur'an. In *Proceedings of the International Conference on Islamic Applications in Computer Science And Technology*, 2015.

[79] H. Halpin, P. Hayes, J. McCusker, D. McGuinness, and H.: Thompson. *When owl:sameAs IsnâÃŕÂ£Â¡ÃŕÂ£Â¡t the Same: An Analysis of Identity in Linked Data, Intl.* Semantic Web Conference, 2010.

[80] Florian Hanika, Gerhard Wohlgenannt, and Marta Sabou. The uComp protégé plugin for crowdsourcing ontology validation. In *Proceedings of the 2014 International Conference on Posters & Demonstrations Track-Volume 1272*, pages 253–256. CEUR-WS.org, 2014.

[81] Florian Hanika, Gerhard Wohlgenannt, and Marta Sabou. The ucomp protÃľgÃľ plugin: Crowdsourcing enabled ontology engineering. In *Knowledge*

Engineering and Knowledge Management, pages 181–196. Springer International Publishing, 2014. ISBN 3319137034.

[82] Jan Hannemann and Jürgen Kett. Linked data for libraries. In *Proc of the world library and information congress of the Int'l Federation of Library Associations and Institutions (IFLA)*.

[83] Fouzi Harrag. Text mining approach for knowledge extraction in sahih al-bukhari. *Computers in Human Behavior*, 30:558–566, 2014. ISSN 07475632. doi: 10.1016/j.chb.2013.06.035. URL `http://ac.els-cdn.com/S0747563213002276/1-s2.0-S0747563213002276-main.pdf?_tid=e017b130-498d-11e4-af4d-00000aab0f02&acdnat=1412183622_a88dc4609595748c0cc1e073d9dc70e9`.

[84] Fouzi Harrag and Aboubekeur Hamdi-Cherif. Uml modeling of text mining in arabic language and application to the prophetic traditions hadiths. pages 11–20, 2007.

[85] Fouzi Harrag, Aboubekeur Hamdi-Cherif, and Eyas El-Qawasmeh. Information retrieval architecture for hadith text mining. *Journal of Digital Information Management*, 6(6), 2008. ISSN 0972-7272.

[86] Fouzi Harrag, Aboubekeur Hamdi-Cherif, and Eyas El-Qawasmeh. Vector space model for arabic information retrieval – application to hadith indexing. In *Applications of Digital Information and Web Technologies, 2008. ICADIWT 2008. First International Conference on the*, pages 107–112. IEEE, 2008.

[87] Fouzi Harrag, Eyas El-Qawasmeh, and Abdul Malik Salman Al-Salman. *Extracting Named Entities from Prophetic Narration Texts (Hadith)*, pages 289–297. Springer, 2011. ISBN 3642221904.

[88] Suhaib Hasan. *An introduction to the science of Hadith.* Al-Quran Society, 1994.

[89] Tom Heath and Christian Bizer. Linked data: Evolving the web into a global data space. *Synthesis lectures on the semantic web: theory and technology*, 1(1):1–136, 2011.

[90] James Hendler, Jeanne Holm, Chris Musialek, and George Thomas. Us government linked open data: semantic. data. gov. *IEEE Intelligent Systems*, 27(3):0025–31, 2012. ISSN 1541-1672. URL `<GotoISI>://WOS:000306450800007`. 974QG Times Cited:6 Cited References Count:5.

[91] Francis Heylighen. Conceptions of a global brain: an historical review. *Evolution: Cosmic, Biological, and Social, eds. Grinin, LE, Carneiro, RL, Korotayev AV, Spier F*, pages 274–289, 2011.

[92] M. Hirth, T. HoÃČÅÿfeld, and P.: Tran-Gia. *Anatomy of a Crowdsourcing Platform - Using the Example of Microworkers.com, 5th Intl.* Conf. on Innovative Mobile and Internet Services in Ubiquitous Computing, 2011.

[93] A. Hogan, J. Umbrich, A. Harth, R. Cyganiak, A. Polleres, and S.: Decker. *An Empirical Survey of Linked Data Conformance.* Journal of Web Semantics, 2012.

[94] A. Hogan, A. Zimmermann, J. Umbrich, A. Polleres, S.: Scalable Decker, and Distributed Methods for Entity Matching. *Consolidation and Disambiguation over Linked Data Corpora.* Journal of Web Semantics, 2012.

[95] Lynne C. Howarth. Frbr and linked data: Connecting frbr and linked data. *Cataloging and Classification Quarterly*, 50(5-7):763–776, 2012. ISSN 0163-9374 1544-4554. doi: 10.1080/01639374.2012.680835. URL `http://www.tandfonline.com/doi/pdf/10.1080/01639374.2012.680835`.

[96] D. Howe, M. Costanzo, and P. Fey. et. al.: Big data: The future of biocuration. *Nature*, 455:47–50, 2008.

[97] J Howe. The rise of crowdsourcing. *Wired magazine*, pages 1–5, 2006.

[98] Jeff Howe. *Crowdsourcing: How the power of the crowd is driving the future of business.* Random House, 2008. ISBN 1905211112.

[99] Anna Huang. Similarity measures for text document clustering. In *Proceedings of the sixth new zealand computer science research student conference (NZCSRSC2008), Christchurch, New Zealand*, pages 49–56, 2008.

[100] et al. Inel, Oana. Crowdtruth: Machine-human computation framework for harnessing disagreement in gathering annotated data. In *The Semantic WebâĂŞISWC 2014*, pages 486–504. Springer, 2014.

[101] Oana Inel, Lora Aroyo, Chris Welty, and Robert-Jan Sips. Domain-independent quality measures for crowd truth disagreement. *Detection, Representation, and Exploitation of Events in the Semantic Web*, page 2, 2013.

[102] Oana Inel et al. Crowdtruth: machine-human computation framework for harnessing disagreement in gathering annotated data. In *The Semantic Web–ISWC 2014*, pages 486–504. Springer, 2014.

[103] Panagiotis G Ipeirotis, Foster Provost, and Jing Wang. Quality management on amazon mechanical turk. In *Proceedings of the ACM SIGKDD workshop on human computation*, pages 64–67. ACM, 2010. ISBN 145030222X.

[104] P. Jain, P. Hitzler, P. Yeh, K. Verma, and A.: Sheth. *Linked Data is Merely More Data.* AAAI Spring Symposium "Linked Data Meets AI", Linked AI, 2010.

[105] Prateek Jain, Pascal Hitzler, Peter Z Yeh, Kunal Verma, and Amit P Sheth. Linked data is merely more data. In *AAAI Spring Symposium: linked data meets artificial intelligence*, volume 11, 2010.

[106] G. Joshi-Tope, M. Gillespie, I. Vastrik, P. D'Eustachio, E. Schmidt, B. de Bono, B. Jassal, G. R. Gopinath, G. R. Wu, and L. Matthews and. *and others: Reactome: a knowledgebase of biological pathways*. Nucleic acids research, Oxford Univ Press, 2005.

[107] Mohammad S Khorsheed and Abdulmohsen O Al-Thubaity. Comparative evaluation of text classification techniques using a large diverse arabic dataset. *Language resources and evaluation*, 47(2):513–538, 2013.

[108] A. Kittur, E. H. Chi, and B. Suh. Crowdsourcing user studies with mechanical turk. *Chi 2008: 26th Annual Chi Conference on Human Factors in*

Computing Systems Vols 1 and 2, Conference Proceedings, pages 453–456, 2008. ISSN 9781605580111.

[109] Shirlee-ann Knight and Janice M. Burn. Developing a framework for assessing information quality on the world wide web. *Informing Science: International Journal of an Emerging Transdiscipline*, 8(5):159–172, 2005. Informing Science Institute.

[110] Magnus Knuth and Harald Sack. Data cleansing consolidation with patchr. *The Semantic Web: ESWC 2014 Satellite Events*, pages 231–235.

[111] Magnus Knuth, Johannes Hercher, and Harald Sack. Collaboratively patching linked data. *CoRR*, abs/1204.2715, 2012. URL `http://arxiv.org/abs/1204.2715`.

[112] Sarath Kumar Kondreddi, Peter Triantafillou, and Gerhard Weikum. Combining information extraction and human computing for crowdsourced knowledge acquisition. In *Data Engineering (ICDE), 2014 IEEE 30th International Conference on*, pages 988–999. IEEE, 2014.

[113] D. Kontokostas, A. Zaveri, S. Auer, and J. Lehmann. Triplecheckmate: A tool for crowdsourcing the quality assessment of linked data. *Knowledge Engineering and the Semantic Web (Kesw 2013)*, 394:265–272, 2013. ISSN 1865-0929.

[114] Markus Krause, Aneta Takhtamysheva, Marion Wittstock, and Rainer Malaka. Frontiers of a paradigm: Exploring human computation with digital

games. In *Proceedings of the ACM SIGKDD Workshop on Human Computation*, HCOMP '10, pages 22–25, New York, NY, USA, 2010. ACM. ISBN 978-1-4503-0222-7.

[115] Anand P Kulkarni, Matthew Can, and Bjoern Hartmann. Turkomatic: automatic recursive task and workflow design for mechanical turk. In *CHI'11 Extended Abstracts on Human Factors in Computing Systems*, pages 2053–2058. ACM, 2011.

[116] Yen-ling Kuo, Jong-Chuan Lee, Kai-yang Chiang, Rex Wang, Edward Shen, Cheng-wei Chan, and Jane Yung-jen Hsu. Community-based game design: experiments on social games for commonsense data collection. In *Proceedings of the ACM SIGKDD Workshop on Human Computation*, pages 15–22. ACM, 2009. ISBN 1605586722.

[117] Jens Lehmann, Tri Quan Nguyen, and Timofey Ermilov. Can we create better links by playing games? In *Semantic Computing (ICSC), 2013 IEEE Seventh International Conference on*, pages 322–329. IEEE, 2013.

[118] Shang-Wen Daniel Li and Piotr Mitros. Learnersourced recommendations for remediation. In *Advanced Learning Technologies (ICALT), 2015 IEEE 15th International Conference on*, pages 411–412. IEEE, 2015.

[119] Henry Lieberman, Dustin Smith, and Alea Teeters. Common consensus: a web-based game for collecting commonsense goals. 2007. ACM Workshop on Common Sense for Intelligent Interfaces.

[120] H. Lin, J. Davis, and Y.: Zhou. Ontological services using crowdsourcing. In *21st Australian Conference on Information Systems*, 2010.

[121] Greg Little, Lydia B Chilton, Max Goldman, and Robert C Miller. Turkit: tools for iterative tasks on mechanical turk. In *Proceedings of the ACM SIGKDD workshop on human computation*, pages 29–30. ACM, 2009.

[122] Greg Little, Lydia B. Chilton, Max Goldman, and Robert C. Miller. Exploring iterative and parallel human computation processes. In *Proceedings of the ACM SIGKDD Workshop on Human Computation*, HCOMP '10, pages 68–76, New York, NY, USA, 2010. ACM. ISBN 978-1-4503-0222-7.

[123] Greg Little, Lydia B Chilton, Max Goldman, and Robert C Miller. Turkit: human computation algorithms on mechanical turk. In *Proceedings of the 23nd annual ACM symposium on User interface software and technology*, pages 57–66. ACM, 2010.

[124] Vanessa Lopez, Andriy Nikolov, Miriam Fernandez, Marta Sabou, Victoria Uren, and Enrico Motta. Merging and ranking answers in the semantic web: The wisdom of crowds. In *The semantic web*, pages 135–152. Springer, 2009.

[125] Nuno Luz, Nuno Silva, and Paulo Novais. Generating human-computer micro-task workflows from domain ontologies. In *Human-Computer Interaction. Theories, Methods, and Tools*, pages 98–109. Springer, 2014.

[126] Nuno Luz, Nuno Silva, and Paulo Novais. A survey of task-oriented crowdsourcing. *Artificial Intelligence Review*, pages 1–27, 2014.

[127] Thomas W Malone, Robert Laubacher, and Chrysanthos Dellarocas. Harnessing crowds: Mapping the genome of collective intelligence. *MIT Sloan Research Paper 4732-09*, 2009.

[128] Thomas W Malone, Robert Laubacher, and Chrysanthos Dellarocas. The collective intelligence genome. *IEEE Engineering Management Review*, 38 (3):38, 2010. ISSN 0360-8581.

[129] A. Marcus, E. Wu, S. Madden, and R.: Crowdsourced Databases: Miller. Query processing with people. In *The Biennial Conference on Innovative Data Systems Research (CIDR*, 2011.

[130] Thomas Markotschi and Johanna Völker. Guesswhat?!–human intelligence for mining linked data. In *Proceedings of the 1st Workshop on Knowledge Injection into and Extraction from Linked Data (KIELD)*. CEUR-WS.org, 2010.

[131] Milan Markovic, Peter Edwards, David Corsar, and Jeff Z Pan. The crowd and the web of linked data: A provenance perspective. In *AAAI Spring Symposium: Wisdom of the Crowd*, 2012.

[132] Milan Markovic, Peter Edwards, David Corsar, and Jeff Z Pan. Demo: Managing the provenance of crowdsourced disruption reports. In *Provenance and Annotation of Data and Processes*, pages 209–213. Springer, 2012. ISBN 3642342213.

[133] Milan Markovic, Peter Edwards, and David Corsar. Utilising provenance

to enhance social computation. In *The Semantic Web–ISWC 2013*, pages 440–447. Springer, 2013. ISBN 3642413374.

[134] Milan Markovic, Peter Edwards, and David Corsar. A role for provenance in social computation. In *Proceedings of the 1st International Conference on Crowdsourcing the Semantic Web - Volume 1030*, CrowdSem'13, pages 93–95, Aachen, Germany, Germany, 2013. CEUR-WS.org.

[135] Milan Markovic, Peter Edwards, and David Corsar. SC-PROV: A provenance vocabulary for social computation. In *Provenance and Annotation of Data and Processes*, pages 285–287. Springer, 2014.

[136] Alistair Miles, Brian Matthews, Michael Wilson, and Dan Brickley. Skos core: Simple knowledge organisation for the web. In *Proceedings of the 2005 International Conference on Dublin Core and Metadata Applications: Vocabularies in Practice*, DCMI '05, pages 1:1–1:9. Dublin Core Metadata Initiative, 2005. ISBN 8489315442, 9788489315440. URL `http://dl.acm.org/citation.cfm?id=1383465.1383467`.

[137] Eric Miller and Micheline Westfall. Linked data and libraries. *The Serials Librarian*, 60(1-4):17–22, 2011.

[138] Patrick Minder and Abraham Bernstein. Crowdlang: A programming language for the systematic exploration of human computation systems. In *Social Informatics*, pages 124–137. Springer, 2012. ISBN 3642353851.

[139] Piotr Mitros. Learnersourcing of complex assessments. In *Proceedings of the*

Second (2015) ACM Conference on Learning@ Scale, pages 317–320. ACM, 2015.

[140] J. M. Mortensen. Crowdsourcing ontology verification. *Semantic Web – ISWC 2013, Part II*, pages 448–455, 2013. ISSN 0302-9743.

[141] J. M. Mortensen, M. A. Musen, and N. F. Noy. Crowdsourcing the verification of relationships in biomedical ontologies. *AMIA Annu Symp Proc*, 2013:1020–9, 2013. ISSN 1942-597X (Electronic) 1559-4076 (Linking).

[142] Jonathan M Mortensen, Paul R Alexander, Mark A Musen, and Natalya F Noy. Crowdsourcing ontology verification.

[143] Jonathan M. Mortensen, Mark A. Musen, and Natalya F. Noy. Developing crowdsourced ontology engineering tasks: An iterative process. In *Proceedings of the 1st International Conference on Crowdsourcing the Semantic Web - Volume 1030*, CrowdSem'13, pages 79–88, Aachen, Germany, Germany, 2013. CEUR-WS.org.

[144] Jonathan M Mortensen, Mark A Musen, and Natalya F Noy. Ontology quality assurance with the crowd. In *First AAAI Conference on Human Computation and Crowdsourcing - Works in Progress and Demonstration Abstracts*, pages 54–55, 2013.

[145] Mohammed Naji Al-Kabi, Ghassan Kanaan, Riyad Al-Shalabi, Saja I Al-Sinjilawi, and Ronza S Al-Mustafa. Al-hadith text classifier. *Journal of Applied Sciences*, 5:584–587, 2005.

[146] Tom Narock and Pascal Hitzler. Crowdsourcing semantics for big data in geoscience applications. In *2013 AAAI Fall Symposium Series*, 2013.

[147] Mathias Niepert, Cameron Buckner, and Colin Allen. Working the crowd: Design principles and early lessons from the social-semantic web. 2009. Proceedings of the Workshop on Web 3.0: Merging Semantic Web and Social Web at ACM Hypertext.

[148] Natalya F. Noy, Jonathan Mortensen, Mark A. Musen, and Paul R. Alexander. Mechanical turk as an ontology engineer?: Using microtasks as a component of an ontology-engineering workflow. In *Proceedings of the 5th Annual ACM Web Science Conference*, WebSci '13, pages 262–271, New York, NY, USA, 2013. ACM. ISBN 978-1-4503-1889-1.

[149] Jasper Oosterman, Alessandro Bozzon, and Geert-Jan et al. Houben. Crowd vs. experts: nichesourcing for knowledge intensive tasks in cultural heritage. pages 567–568. Int. WWW Conferences Steering Committee, 2014.

[150] Jasper Oosterman, Archana Nottamkandath, Chris Dijkshoorn, Alessandro Bozzon, Geert-Jan Houben, and Lora Aroyo. Crowdsourcing knowledge-intensive tasks in cultural heritage. In *Proceedings of the 2014 ACM Conference on Web Science*, WebSci '14, pages 267–268, New York, NY, USA, 2014. ACM. ISBN 978-1-4503-2622-3.

[151] A. Parameswaran and N.: Polyzotis. *Answering Queries using Humans*. Algorithms and Databases, Conf. on Innovative Data Systems Research, 2011.

[152] Ei Pa Pa Pe-Than, Dion Hoe-Lian Goh, and Chei Sian Lee. A typology of human computation games: an analysis and a review of current games. *Behaviour & Information Technology*, 34(8):809–824, 2015.

[153] A Peres, M Lopes, and O Corcho. *Ontological Engineering: with Examples from the Areas of knowledge Management, e-Commerce and Semantic Web*. Springer, 2004.

[154] Abu Ameenah Bilal Philips. *Usool at-Tafseer: The Methodology of Qur'aanic Explanation.* AS Noordeen, 2002.

[155] Nelson Piedra, Edmundo Tovar, Ricardo Colomo-Palacios, Jorge Lopez-Vargas, and Janneth Alexandra Chicaiza. Consuming and producing linked open data: the case of opencourseware. *Program: electronic library and information systems*, 48(1):16–40, 2014. ISSN 0033-0337. doi: 10.1108/prog-07-2012-0045.

[156] Alexander J. Quinn and Benjamin B. Bederson. Human computation: A survey and taxonomy of a growing field. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, CHI'11, pages 1403–1412, New York, NY, USA, 2011. ACM. ISBN 978-1-4503-0228-9.

[157] Yehya M. Dalloul Rebhi S. Baraka. Building hadith ontology to support the authenticity of isnad. *International Journal on Islamic Applications in Computer Science And Technology*, 2(1):25–39, 2014.

[158] Daniela Retelny, Sébastien Robaszkiewicz, Alexandra To, Walter S Lasecki,

Jay Patel, Negar Rahmati, Tulsee Doshi, Melissa Valentine, and Michael S Bernstein. Expert crowdsourcing with flash teams. In *Proceedings of the 27th annual ACM symposium on User interface software and technology*, pages 75–85. ACM, 2014.

[159] Antonio J. Roa-Valverde. Combining gamification, crowdsourcing and semantics for leveraging linguistic open data. In *Proceedings of the Linked Learning meets LinkedUp Workshop: Learning and Education with the Web of Data (LILE 2014)*. CEUR-WS.org, 2014.

[160] M. Sabou, A. Scharl, and M. Fols. Crowdsourced knowledge acquisition: Towards hybrid-genre workflows. *International Journal on Semantic Web and Information Systems*, 9(3):14–41, 2013. ISSN 1552-6283.

[161] Marta Sabou, Kalina Bontcheva, and Arno Scharl. Crowdsourcing research opportunities: lessons from natural language processing. In *Proceedings of the 12th International Conference on Knowledge Management and Knowledge Technologies*, page 17. ACM, 2012.

[162] Marta Sabou, Arno Scharl, and Föls Michael. Crowdsourced knowledge acquisition: Towards hybrid-genre workflows. *International Journal on Semantic Web and Information Systems*, 9(3):14–41, 2013.

[163] C. Sarasua, E. Simperl, and N. F.: Crowdmap: Noy. Crowdsourcing ontology alignment with microtasks. *ISWC*, 2012, 2012.

[164] Cristina Sarasua, Elena Simperl, and Natalya F Noy. Crowdmap: Crowd-

sourcing ontology alignment with microtasks. In *The Semantic Web–ISWC 2012*, pages 525–541. Springer, 2012.

[165] Arno Scharl, Marta Sabou, and Michael Föls. Climate quiz: A web application for eliciting and validating knowledge from social networks. In *Proceedings of the 18th Brazilian Symposium on Multimedia and the Web*, WebMedia '12, pages 189–192, New York, NY, USA, 2012. ACM. ISBN 978-1-4503-1706-1.

[166] Nigel Shadbolt, Kieron O'Hara, Tim Berners-Lee, Nicholas Gibbins, Hugh Glaser, and Wendy Hall. Linked open government data: Lessons from data. gov. uk. *IEEE Intelligent Systems*, 27(3):16–24, 2012. ISSN 1541-1672. URL `<GotoISI>://WOS:000306450800006`. 974QG Times Cited:6 Cited References Count:10.

[167] Abdul-Baquee M Sharaf and Eric Atwell. Qursim: A corpus for evaluation of relatedness in short texts. In *LREC*, pages 2295–2302, 2012.

[168] Mohamed Ahmed Sherif and Axel-Cyrille Ngonga Ngomo. Semantic Quran - a multilingual resource for natural-language processing. *Semantic Web*, 6 (4):339–345, 2015.

[169] Mohamed Ahmed Sherif and Axel-Cyrille Ngonga Ngomo. Semantic quran: A multilingual resource for natural-language processing. *Semantic Web*, 2014. ISSN 1570-0844.

[170] A. Sheth:. Computing for Human Experience. URL `http://wiki.knoesis.org/index.php/Computing_For_Human_Experience`.

[171] Jakub Šimko. Harnessing manpower for creating semantics. *Information Sciences & Technologies: Bulletin of the ACM Slovakia*, 5(3), 2013. Citeseer.

[172] Jakub Šimko and Mária Bieliková. Little search game: Lightweight domain modeling. In *Semantic Acquisition Games*, pages 51–65. Springer, 2014.

[173] Jakub Simko and MÃąria BielikovÃą. Semantic acquisition games. pages 35–50, 2014. ISSN 978-3-319-06114-6.

[174] Jakub Šimko, Michal Tvarozek, and Mária Bieliková. Little search game: Term network acquisition via a human computation game. In *Proceedings of the 22Nd ACM Conference on Hypertext and Hypermedia*, HT '11, pages 57–62, New York, NY, USA, 2011. ACM. ISBN 978-1-4503-0256-2.

[175] E. Simperl, M. Acosta, and B.: Norton. *A Semantically Enabled Architecture for Crowdsourced Linked Data Management, Proc. of the 1st Intl.* Workshop on Crowdsourcing, 2012.

[176] Elana Simperl, Stephan Wölger, Stefan Thaler, and Katharina Siorpaes. *Perspectives on Ontology Learning*, volume 18, chapter Learning ontologies via games with a purpose. IOS Press, 2014.

[177] Elena Simperl. Crowdsourcing semantic data management: Challenges and opportunities. In *Proceedings of the 2Nd International Conference on Web*

Intelligence, Mining and Semantics, WIMS '12, pages 1:1–1:3, New York, NY, USA, 2012. ACM. ISBN 978-1-4503-0915-8.

[178] Elena Simperl and Markus Luczak-Rösch. Collaborative ontology engineering: a survey. *The Knowledge Engineering Review*, 29(01):101–131, 2014. Cambridge Univ Press.

[179] Elena Simperl, Barry Norton, and Denny Vrandecic. Crowdsourcing tasks in linked data management. In *Proceeding of 2nd workshop on consuming Linked Data, COLD2011 (co-located with the 10th International Semantic Web Conference ISWC)*, volume 782. CEUR-WS.org, 2011.

[180] Elena Simperl, Maribel Acosta, and Barry Norton. A semantically enabled architecture for crowdsourced linked data management. In *CrowdSearch*, pages 9–14. Citeseer, 2012.

[181] Elena Simperl, Stephan Wölger, Stefan Thaler, Barry Norton, and Tobias Bi̊rger. Combining human and computation intelligence: the case of data interlinking tools. *International Journal of Metadata, Semantics and Ontologies*, 7(2):77–92, 2012. ISSN 1744-2621.

[182] Elena Simperl, Maribel Acosta, and Fabian Flöck. Knowledge engineering via human computation. In *Handbook of Human Computation*, pages 131–151. Springer, 2013.

[183] Katharina Siorpaes and Martin Hepp. Games with a purpose for the semantic web. *IEEE Intelligent Systems*, 23(3):50–60, 2008. ISSN 1541-1672.

[184] Katharina Siorpaes and Martin Hepp. *Ontogame: Weaving the semantic web by online games.* Springer, 2008. ISBN 3540682333.

[185] Katharina Siorpaes and Elena Simperl. Incentives, motivation, participation, games: Human computation for linked data. In *Proceedings of the Workshop on Linked Data in the Future Internet at the Future Internet Assembly*, volume 700. CEUR-WS.org, 2010.

[186] Katharina Siorpaes and Elena Simperl. Human intelligence in the process of semantic content creation. *World Wide Web*, 13(1-2):33–59, 2010. ISSN 1386-145X.

[187] Katharina Siorpaes, Martin Hepp, Andreas Klotz, and Michael Hackl. myontology: tapping the wisdom of crowds for building ontologies. Citeseer, 2008.

[188] Guillermo Soberón, Lora Aroyo, Chris Welty, Oana Inel, Hui Lin, and Manfred Overmeen. Measuring crowd truth: Disagreement metrics combined with worker behavior filters. In *Proc. of CrowdSem 2013 Workshop, ISWC*, 2013.

[189] Ge Song, Yunming Ye, Xiaolin Du, Xiaohui Huang, and Shifu Bie. Short text classification: A survey. *Journal of Multimedia*, 9(5):635–643, 2014.

[190] Alexander Strehl, Joydeep Ghosh, and Raymond Mooney. Impact of similarity measures on web-page clustering. In *Workshop on Artificial Intelligence for Web Search (AAAI 2000)*, pages 58–64, 2000.

[191] Stefan Thaler, Elena Simperl, and Katharina Siorpaes. Spotthelink: playful alignment of ontologies. In *Proceedings of the 2011 ACM Symposium on Applied Computing*, pages 1711–1712. ACM, 2011. ISBN 1450301134.

[192] Stefan Thaler, Elena Paslaru Bontas Simperl, and Katharina Siorpaes. Spotthelink: A game for ontology alignment. *Wissensmanagement*, 182:246–253, 2011.

[193] Stefan Thaler, Katharina Siorpaes, David Mear, Elena Simperl, and Carl Goodman. Seafish: a game for collaborative and visual image annotation and interlinking. In *The Semanic Web: Research and Applications*, pages 466–470. Springer, 2011.

[194] Stefan Thaler, Elena Simperl, and Stephan Wolger. An experiment in comparing human-computation techniques. *Internet Computing, IEEE*, 16(5): 52–58, 2012. ISSN 1089-7801.

[195] Andreas Thalhammer, Magnus Knuth, and Harald Sack. Evaluating entity summarization using a game-based ground truth. In *The Semantic Web–ISWC 2012*, pages 350–361. Springer, 2012. ISBN 3642351727.

[196] David Vickrey, Aaron Bronzan, William Choi, Aman Kumar, Jason Turner-Maier, Arthur Wang, and Daphne Koller. Online word games for semantic data collection. In *Proceedings of the Conference on Empirical Methods in Natural Language Processing*, EMNLP '08, pages 533–542, Stroudsburg, PA, USA, 2008. Association for Computational Linguistics.

[197] Duc-Thuan Vo and Cheol-Young Ock. Learning to classify short text from scientific documents using topic models with various types of knowledge. *Expert Systems with Applications*, 42(3):1684–1698, 2015.

[198] J. Volz, C. Bizer, M. Gaedke, and G.: Kobilarov. Silk âÃŕÂ£Â¡“ a link discovery framework for the web of data. 2, 2009.

[199] Luis Von Ahn. Games with a purpose. *Computer*, 39(6):92–94, 2006. ISSN 0018-9162.

[200] Luis Von Ahn. Human computation. In *Design Automation Conference, 2009. DAC'09. 46th ACM/IEEE*, pages 418–419. IEEE, 2009. ISBN 1605584975.

[201] Luis Von Ahn and Laura Dabbish. Labeling images with a computer game. In *Proceedings of the SIGCHI conference on Human factors in computing systems*, pages 319–326. ACM, 2004. ISBN 1581137028.

[202] Luis Von Ahn and Laura Dabbish. Designing games with a purpose. *Communications of the ACM*, 51(8):58–67, 2008. ISSN 0001-0782.

[203] Luis von Ahn, Mihir Kedia, and Manuel Blum. Verbosity: A game for collecting common-sense facts. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, CHI '06, pages 75–78, New York, NY, USA, 2006. ACM. ISBN 1-59593-372-7.

[204] Jakub Šimko, Michal Tvarožek, and Mária Bieliková. Semantics discovery

via human computation games. *International Journal on Semantic Web & Information Systems*, 7(3):23–45, July 2011. ISSN 1552-6283.

[205] Jörg Waitelonis, Nadine Ludwig, Magnus Knuth, and Harald Sack. Who-knows? evaluating linked data heuristics with a quiz that cleans up dbpedia. *Interactive Technology and Smart Education*, 8(4):236–248, 2011. Emerald Group Publishing Limited.

[206] Aobo Wang, Cong Duy Vu Hoang, and Min-Yen Kan. Perspectives on crowd-sourcing annotations for natural language processing. *Language resources and evaluation*, 47(1):9–31, 2013.

[207] J. Wang, T. Kraska, M. Franklin, and J.: CrowdER: Crowdsourcing Entity Resolution Feng. Proc. of the vldb endowment. *Vol.*, 5:10, 2012.

[208] Stuart Weibel, John Kunze, Carl Lagoze, and Misha Wolf. Dublin core metadata for resource discovery. Technical report, 1998.

[209] Sarah Weir, Juho Kim, Krzysztof Z Gajos, and Robert C Miller. Learner-sourcing subgoal labels for how-to videos. In *Proceedings of the 18th ACM Conference on Computer Supported Cooperative Work & Social Computing*, pages 405–416. ACM, 2015.

[210] Lina Wolf, Magnus Knuth, Johannes Osterhoff, and Harald Sack. RISQ! renowned individuals semantic quiz: A jeopardy like quiz game for ranking facts. In *Proceedings of the 7th International Conference on Semantic Sys-*

tems, I-Semantics'11, pages 71–78, New York, NY, USA, 2011. ACM. ISBN 978-1-4503-0621-8.

[211] Aliyu Rufai Yauri, Rabiah Abdul Kadir, Azreen Azman, and Masrah Azrifah Azmi Murad. Quranic-based concepts: Verse relations extraction using manchester owl syntax. In *Information Retrieval and Knowledge Management (CAMP), 2012 International Conference on*, pages 317–321. IEEE. ISBN 1467310913.

[212] Aliyu Rufai Yauri, Rabiah Abdul Kadir, Azreen Azman, and Masrah Azrifah Azmi Murad. Quranic verse extraction base on concepts using owl-dl ontology. *Research Journal of Applied Sciences, Engineering and Technology*, 6(23):4492–4498, 2013.

[213] Chih-Hao Yu, Tudor Groza, and Jane Hunter. Reasoning on crowd-sourced semantic annotations to facilitate cataloguing of 3d artefacts in the cultural heritage domain. In *The Semantic Web–ISWC 2013*, pages 228–243. Springer, 2013. ISBN 3642413374.

[214] M. Yuen. *et al.: A Survey of Crowdsourcing Systems, IEEE Intl.* Conf. on Social Comp, 2011.

[215] M. Yuen, I. King, and K.: Leung. Task matching in crowdsourcing, proc. of the 4th ieee intl. *Conf. on Internet of Things, and Cyber, Physical and Social Comp. (CPSComÃČÂŋ*, 11, 2011.

[216] Man-Ching Yuen, Irwin King, and Kwong-Sak Leung. A survey of crowd-

sourcing systems. In *Privacy, Security, Risk and Trust (PASSAT) and 2011 IEEE Third Inernational Conference on Social Computing (SocialCom), 2011 IEEE Third International Conference on*, pages 766–773. IEEE, 2011.

[217] Man-Ching Yuen, Irwin King, and Kwong-Sak Leung. Task matching in crowdsourcing. In *Internet of Things (iThings/CPSCom), 2011 International Conference on and 4th International Conference on Cyber, Physical and Social Computing*, pages 409–412. IEEE, 2011.

[218] Man-Ching Yuen, Irwin King, and Kwong-Sak Leung. Taskrec: A task recommendation framework in crowdsourcing systems. *Neural Processing Letters*, pages 1–16, 2014.

[219] Amrapali Zaveri, Anisa Rula, Andrea Maurino, Ricardo Pietrobon, Jens Lehmann, Sören Auer, and Pascal Hitzler. Quality assessment methodologies for linked open data. *Semantic Web Journal*, 7(1):63–93, 2015.

قُلِ ٱللَّهُمَّ مَٰلِكَ ٱلْمُلْكِ تُؤْتِى ٱلْمُلْكَ مَن تَشَآءُ وَتَنزِعُ ٱلْمُلْكَ مِمَّن تَشَآءُ
وَتُعِزُّ مَن تَشَآءُ وَتُذِلُّ مَن تَشَآءُ ۖ بِيَدِكَ ٱلْخَيْرُ ۖ إِنَّكَ عَلَىٰ كُلِّ شَىْءٍ قَدِيرٌ ۝
تُولِجُ ٱلَّيْلَ فِى ٱلنَّهَارِ وَتُولِجُ ٱلنَّهَارَ فِى ٱلَّيْلِ ۖ وَتُخْرِجُ ٱلْحَىَّ مِنَ ٱلْمَيِّتِ
وَتُخْرِجُ ٱلْمَيِّتَ مِنَ ٱلْحَىِّ ۖ وَتَرْزُقُ مَن تَشَآءُ بِغَيْرِ حِسَابٍ ۝

Say, "O Allah , Owner of Sovereignty, You give sovereignty to whom You will and You take sovereignty away from whom You will. You honor whom You will and You humble whom You will. In Your hand is [all] good. Indeed, You are over all things competent.You cause the night to enter the day, and You cause the day to enter the night; and You bring the living out of the dead, and You bring the dead out of the living. And You give provision to whom You will without account."

— *[Al-Quran, Ale-Imraan,3:26-27]*

رَبَّنَا تَقَبَّلْ مِنَّآ إِنَّكَ أَنتَ ٱلسَّمِيعُ ٱلْعَلِيمُ
وَتُبْ عَلَيْنَآ إِنَّكَ أَنتَ ٱلتَّوَّابُ ٱلرَّحِيمُ