Aemoo* : exploring knowledge on the Web

Andrea Giovanni Nuzzolese
DISI, University of Bologna/STLab ISTC-CNR
nuzzoles@cs.unibo.it

Valentina Presutti
STLab ISTC-CNR
valentina.presutti@cnr.it

Aldo Gangemi
LIPN, Paris 13/STLab ISTC-CNR
aldo.gangemi@lipn.univ-paris13.fr

Alberto Musetti
STLab ISTC-CNR
lleipuner@researchlabs.org

Paolo Ciancarini
DISI, University of Bologna/STLab ISTC-CNR
ciancarini@cs.unibo.it

ABSTRACT
Aemoo is a Semantic Web application supporting knowledge exploration on the Web. Through a keyword-based search interface, users can gather an effective summary of the knowledge about an entity, according to Wikipedia, Twitter, and Google News. Summaries are designed by applying lenses based on a set of empirically discovered knowledge patterns.

EXPLORING WEB KNOWLEDGE
The Web is a huge source of knowledge and one of the main research challenges is to make such knowledge easily and effectively accessible to Web users. Applications from the Web of Data, social networks, news services, search engines, etc., attempt to address this requirement, which is still far from being solved due to the many challenges arising, e.g. from the heterogeneity of sources, different representations, implicit semantics of links, etc. Existing semantic mashup and browsing applications, such as [5, 2, 3], mostly focus on presenting linked data coming from different sources, and visualizing it in interfaces that mirror the linked data structure. Typically, they rely on semantic relations that are explicitly asserted in the linked datasets, or in explicit annotations, e.g., microdata, without exploiting additional knowledge, e.g. coming from hypertext links, which limits the data provided, its visualization, and navigation. In practice, the problem of delivering tailored and contextualized knowledge remains unsolved, since retrieved knowledge is returned without tailoring it to any context-based rationale.

Other applications focus on text enrichment by performing identity resolution of named entities. Examples are: Zemanta[1], Stanbol Enhancer[2] and Calais[3]. Such applications are useful for enhancing text with hypertext links to related Web pages and pictures, and sometimes they provide information about the type of the identified entities, which can be useful for designing simple faceted interfaces. However, their approach does not seem inspired by relevance rationales, e.g. their results are provided without any explanation or criterion of why a piece of news, or a set of resources, is to be considered relevant.

To the best of our knowledge no existing approach attempts to organize or filter knowledge before presenting it by drawing a meaningful boundary around the retrieved data in order to limit the visualized results to what is meaningful/useful. In this paper, we present **Aemoo**[4], a Semantic Web application supporting knowledge exploration on the Web, based on knowledge patterns [1]. Aemoo exploits Semantic Web technologies as well as the structure of hypertext links for enriching query results with relevant knowledge coming from diverse Web sources. In addition, Aemoo organizes and filters the retrieved knowledge in order to show only relevant information to users, and provides a motivation why a certain piece of information is included.

The rest of the paper is organize as follows: after describing three simple use case scenarios for Aemoo in next section, we describe our approach and provide details about Aemoo architecture and the technologies used for its implementation. Finally, we conclude and discuss ongoing work.

USAGE SCENARIOS AND AEMOO INTERFACE
In this section we describe Aemoo functionalities through three simple scenarios.

Scenario 1: Knowledge Aggregation and Explanations.
Pedro is a high school student, his homework is to write a report about Immanuel Kant (IK). He types "Immanuel Kant" in the search interface, and gets a summary page about him (cf. Figure 1). On the left side of the page, Pedro reads that IK is a philosopher, together with some general information

*http://www.aemoo.org

[1]http://www.zemanta.com/
[2]http://stanbol.apache.org
[3]http://www.opencalais.com/
[4]Aemoo, http://www.aemoo.org

WebSci'13, May 1 – May 5, 2013, Paris, France.

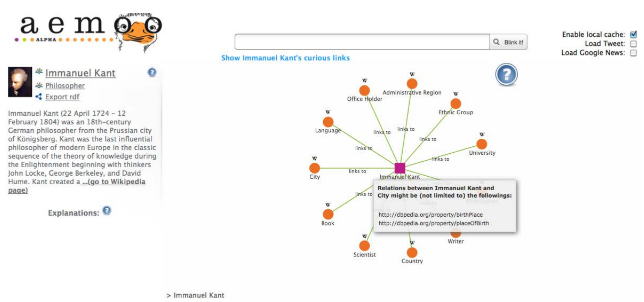

Figure 1. Aemoo: initial summary page for query "Immanuel Kant".

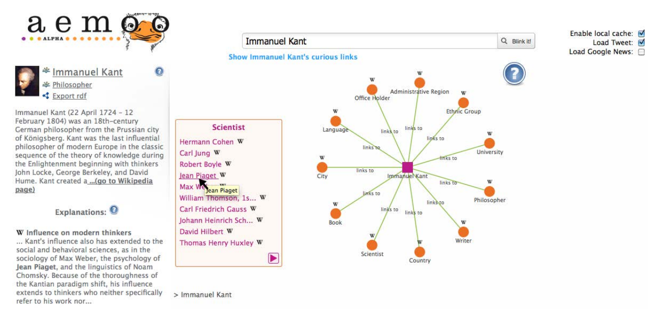

Figure 2. Aemoo: browsing relations between "Immanuel Kant" and scientists.

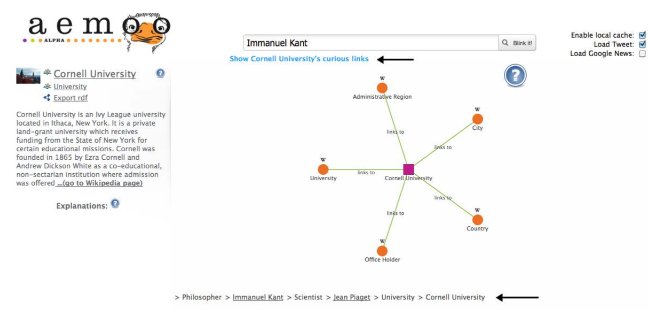

Figure 3. Aemoo: breadcrumb and curiosity

about him, and a thumbnail image. This information will be enriched as a consequence of, and during, his navigation. At the same time, a concept map built around IK (as a central squared node) appears at the center-right of the page. Nodes represent sets of resources of a certain type (the node type is shown as a label), we refer to them as *set nodes*. Additionally, icons on a set node indicate the source(s) from which its contained information is taken, e.g. Wikipedia. Set nodes change depending on the type of the subject (Philosopher in this case), according to the "knowledge pattern" associated with such type (see next section). According to an empirical study described in [4] these set nodes are the types of things that a user would intuitively expect to see in a summary description of a Philosopher, .

An infotip appearing when hovering over a link between IK and a set node, shows a list of possible semantic relations explaining that specific link type, according to their frequencies in DBpedia (cf. Figure 1). Such list is not exhaustive, it only shows existing DBpedia relations (extracted from Wikipedia infoboxes) between two specific types. For example, the relations between IK and cities could be `birthPlace` or `placeOfBirth` if we considered only DBpedia asserted relations. This example shows the limit of the current DBpedia relations coverage, in fact cities can be related to Immanuel Kant also for other reasons, which are explained by Aemoo in the explanation section (left-bottom side of the interface).

IK links to a set of scientists, which is interesting information for our user Pedro, who wants to know more about this relation. By hovering on a set node e.g., *Scientist*, he triggers the visualization of a list of resources contained in this set, meaning that those resources are connected to IK (cf. Figure 2. By hovering on a specific entity of the set e.g., Jean Piaget, new information is visualized under the "Explanations" section (left-bottom). Such information explains the meaning of that connection. In the example, Jean Piaget is linked to IK because his work was influenced by Kant's.

Explanations come from different possible sources i.e., Wikipedia, Twitter, and Google news. The sources to be used can be selected by users through a set of checkbox put in the top-right corner of the interface.

Scenario 2: Exploratory search.

Pedro would like to collect some more information about Jean Piaget, hence, he clicks on that entity in the list. Aemoo changes context from IK to Jean Piaget, showing a new summary page for this scientist. Pedro can perform exploratory search by inspecting set nodes and lists associated with Jean Piaget, and possibly other entities. Figure 3 shows the situation after some exploration steps. Through the breadcrumb (located at the center-bottom of the interface) Pedro can go back and forth, and revisit his exploration path and its associated knowledge.

Scenario 3: Curiosity.

Eva is an editorial board member of a TV program that dedicates each episode to a different country. She has to edit an episode about Italy. She uses Aemoo, as described above, for building a summary about the country that can be useful for the introductory part of the show. However, she wants to find peculiar information that can make the episode more interesting to her audience. Aemoo helps her on this task through the "curiosity" functionality, that can be triggered by clicking on the link between the search field and the concept map (cf. Figure 3). Aemoo will change perspective and will provide a new summary for the same entity. Eva will be presented with additional knowledge about Italy, which was not previously included in the summary. What is now shown are "special" facts about Italy, things that are not commonly used to describe a country. Knowledge is again visualized as a concept map, and enriched with news and tweets just as it happened for the previous summary, but this time the set nodes are selected with a different criterion: they are types of resources that are unusual to be included in the description of a country, hence possibly peculiar.

SUMMARIZING KNOWLEDGE ABOUT ENTITIES

This section describes how Aemoo exploits semantics to perform its task and to improve the user experience.

Identity resolution

Aemoo exploits DBpedia for identity resolution and to gather Wikipedia knowledge about an entity. This task is performed in two main situations: (1) when a user types a query; (2) when collecting and filtering relevant knowledge.

User queries are processed and matched against DBpedia entities (1) for identifying the identity of the resource referred to in a query. As a result users are provided with a list of possible options (autocompletion), among which they can select the "subject" their exploration (the selected entity is called hereon "subject").

Aemoo currently uses three main sources: Wikipedia, Google News, and Twitter. All entities that are linked to from the Wikipedia page of a subject are classified into set nodes. Additionally, the current stream of Twitter messages and available articles provided by Google News are analyzed in order to identify possible entities that co-occur with mentions of the subject. For example, consider a user that selects "Steve Jobs" as a subject, and the following tweet: "Steve Jobs leaving his place at Apple to Tim Cook". Aemoo would resolve the identity of "Tim Cook" and "Apple" and would add them to the appropriate set nodes of entities related to Steve Jobs.

Entity types and Encyclopedic Knowledge Patterns

Aemoo retrieves the types of the resolved entities, according to the DBpedia taxonomy of types. This type is used for providing users with additional information about the subject (it is indicated on the top-left), and as a criterion for assigning an entity to a certain set node. Types are also used as a criterion for filtering the knowledge to be presented. Aemoo approach is based on the application of lenses based on *Encyclopedic Knowledge Patterns* (EKPs) used as a means for building a knowledge summary about a subject.

Knowledge Patterns are formal conceptual structures that represent the core elements contributing to express knowledge about specific things and contexts, typically event or situation types. They provide a criterion for drawing boundaries around a set of unordered data so as to give them a unique intensional meaning. Knowledge patterns are the computational counterparts of cognitive schemas, used by humans for organizing their knowledge, and for interpreting and making sense of observed facts. For example, if a human looks at a picture showing a person laying on a floor in a room with a chair and a table, and another person standing close to the first one and holding a bloodstained knife, (s)he would interpret it as a homicide scene, would assign the role of victim to the first person, the role of killer to the second person, the role of arm to the knife, and would discard from his/her consideration all the other objects i.e. the chair and the table. In other words, we as humans have the ability to filter relevant knowledge i.e., drawing semantic boundaries, out of a number of facts, and use it for formulating a semantic interpretation. In a similar way, knowledge patterns can be used by machines to filter and make sense of data.

EKPs are a special type of knowledge patterns: they express the core elements that are used for describing entities of a certain type with an encyclopedic task in mind. In a previous work [4] we have empirically extracted about 200 EKPs from Wikipedia, and have demonstrated that they provide cognitively-sound organizational structures of descriptive knowledge. The extracted EKPs are available online[5] as OWL2 ontologies. Aemoo builds an entity summary by applying lenses based on EKPs on data associated with such entity: the concept map of a subject is built by only including the elements i.e., types, of the EKP associated with that subject type. For example, given the subject "Paris", which has type `dbpo:City`[6], and the EKP associated with `dbpo:City` that includes the types: `dbpo:City`, `dbpo:Country`, `dbpo:University`, and `dbpo:OfficeHolder`, Aemoo summary for Paris shows a concept map including a set node for each of these types, and each set containing entities of that type that are linked to from the Wikipedia page of Paris. EKPs are also used for identifying "curiosities" about a subject. Aemoo uses long-tail links - that are normally taken out by a EKP lens - for building a different perspective over the knowledge related to an entity, which includes peculiar facts instead of core knowledge.

Explanations and semantics of links

All entities included in a subject summary are related to it by a hypertext link, or because they co-occur with the subject in a news article or a tweet. The meaning of these relations is explained by the text surrounding the anchor or the co-occurrence reference. Aemoo exploits this aspect by extracting such pieces of text and showing them in association with each specific link.

In summary, Aemoo performs KP-based knowledge exploration, which makes it especially novel. It exploits the structure of linked data, and organizes it by means of EKPs for supporting exploratory search. The use of EKPs allows Aemoo to draw meaningful boundaries around data. In this way, Aemoo performs both *enrichment* and *filtering* of information, based on the structure of EKPs, which reflects the most common way to describe entities of a particular type. Users are guided through their navigation: instead of being presented with a bunch of triples or a big unorganized graph they navigate through *units of knowledge* and move from one to the other without loosing the overview of an entity.

IMPLEMENTATION DETAILS

Aemoo is released as a web application: it consists of a server side component implemented as a Java-based REST service, and a client side component based on HTML and JavaScript. The client side interacts with third party components via REST interfaces through AJAX.

The server side exposes a REST service for retrieving EKP-based graphs as well as "curiosity graphs" about entities. Its input is an entity URI, e.g. dbpedia:Barack_Obama[7]. Its

[5] http://www.ontologydesignpatterns.org/ekp/

[6] dbpo: stands for http:/dbpedia.org/ontology

[7] dbpedia: stands for http://dbpedia.org/resource/

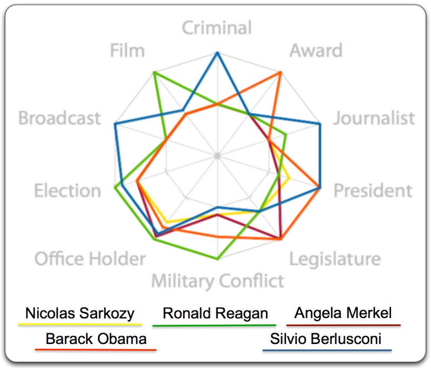

Figure 4. Comparing entities of the same type.

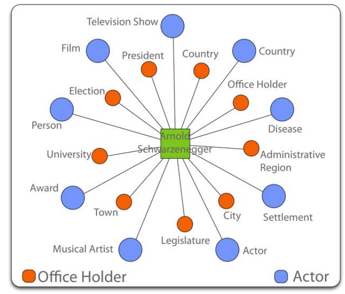

Figure 5. Filtering and visualizing knowledge about an entity by applying different knowledge patterns.

output is an RDF graph corresponding to the summarization based on the EKP associated with the entity type. The RDF graph is obtained by automatically generating a SPARQL CONSTRUCT from the selected EKP.

The client side component handles the graphical visualization of Aemoo through the JavaScript InfoVis Toolkit[8], a Javascript library that supports the creation of interactive data visualizations for the Web. Abstracts and thumbnails are retrieved by querying the DBpedia SPARQL endpoint exposed as a REST service.

Aemoo also detects relations between the inspected entity and other entities from Twitter [9] as well as from Google News[10]. Tweets and Google news are retrieved using their respective REST services. For performing identity resolution on user queries, tweets, and news we use Apache Stanbol Enhancer[11]. Entities recognized in tweets and news are dynamically added to the graph map. Explanations are extracted from the text surrounding wikilinks in the subject Wikipedia page, the text of tweets and of Google news, and are associated with provenance information.

CONCLUSION AND ONGOING WORK

We have presented Aemoo, a Semantic Web application supporting entity summarization and knowledge exploration on the Web based on knowledge patterns.

Currently, we are working on several extensions and we are conducting a user-study for improving its graphical user interface. A number of user-based experiment sessions have been already performed, they include analysis of usability as well as task-based comparison to other tools.

Example of extensions we are currently working on include: (i) empowering the automatic interpretation of hypertext links by using NLP techniques (specially frame detection), combined with knowledge patterns and linked data analysis; (ii) providing visual analytics interfaces that compare different entities having the same type (a mockup is depicted in Figure 4; (iii) providing different views on the same entity by allowing users to change the applied lenses i.e. EKP (a mockup is depicted in Figure 5); (iv) adding a basket functionality, allowing users to save the summary data of their exploration in RDF; and (v) allowing users to edit relation labels in order to make link semantics emerge out of crowd sourcing.

[8] http://thejit.org/
[9] https://search.twitter.com/search.json
[10] https://ajax.googleapis.com/ajax/services/search/news
[11] http://stanbol.apache.org/

REFERENCES

1. A. Gangemi and V. Presutti. Towards a pattern science for the semantic web. *Semantic Web*, 1(1-2):61–68, 2010.
2. P. Heim, S. Hellmann, J. Lehmann, S. Lohmann, and T. Stegemann. RelFinder: Revealing relationships in RDF knowledge bases. In *Proceedings of the 3rd International Conference on Semantic and Media Technologies (SAMT)*, volume 5887 of *Lecture Notes in Computer Science*, pages 182–187. Springer, 2009.
3. P. Heim, J. Ziegler, and S. Lohmann. gFacet: A browser for the web of data. In S. Auer, S. Dietzold, S. Lohmann, and J. Ziegler, editors, *Proceedings of the International Workshop on Interacting with Multimedia Content in the Social Semantic Web (IMC-SSW'08)*, pages 49–58. CEUR-WS, 2008.
4. A. G. Nuzzolese, A. Gangemi, V. Presutti, and P. Ciancarini. Encyclopedic knowledge patterns from wikipedia links. In L. Aroyo, N. Noy, and C. Welty, editors, *Proceedings fo the 10th International Semantic Web Conference (ISWC2011)*, pages 520–536. Springer, 2011.
5. G. Tummarello, R. Cyganiak, M. Catasta, S. Danielczyk, R. Delbru, and S. Decker. Sig.ma: Live views on the web of data. *Web Semantics: Science, Services and Agents on the World Wide Web*, 8(4):355 – 364, 2010. Semantic Web Challenge 2009, User Interaction in Semantic Web research.