

INSA LYON  
UNIVERSITÄT PASSAU

MASTER THESIS

---

# Image Annotation Network

---

*Author:*  
Mael OGIER

*Supervisors:*  
Dr. David COQUIL  
Dr. Elöd EGYED-ZSIGMOND

*This thesis submitted in fulfilment of the requirements  
for the degree of Master of Science*

*in the*

Informatique - Information und Kommunikation (IFIK)  
Double Master Program

August 2015

*"I have a dream for the Web [in which computers] become capable of analyzing all the data on the Web—the content, links, and transactions between people and computers. A "Semantic Web", which should make this possible, has yet to emerge, but when it does, the day-to-day mechanisms of trade, bureaucracy and our daily lives will be handled by machines talking to machines."*

Tim Berners-Lee

# *Acknowledgements*

This master thesis was done in the context of the double master degree Informatique - Information und Kommunikation (IFIK), which brings together two Master programs : a degree in computer engineering at the National Institute of Applied Sciences in Lyon (INSA Lyon) and a Master in Informatik (Schwerpunkt : Information und Kommunikationssysteme) at the University of Passau.

**Add some thanks here : Pr. Kosh, Brunie, Dr. Coquil, Egyed**

I would also like to thanks the Dropkick Murphys for their energizing music and the local producers of coffee worldwide for this magical beverage.

INSA LYON  
UNIVERSITÄT PASSAU

# *Abstract*

IFIK  
Double Master Program

Master of Science

**Image Annotation Network**

by Mael OGIER

The Thesis Abstract is written here (and usually kept to just this page). The page is kept centered vertically so can expand into the blank space above the title too. . .

# Contents

<b>Acknowledgements</b>	<b>ii</b>
<b>Abstract</b>	<b>iii</b>
<b>Contents</b>	<b>iv</b>
<b>List of Figures</b>	<b>vi</b>
<b>List of Tables</b>	<b>vii</b>
<b>Abbreviations</b>	<b>viii</b>
<b>1 Introduction</b>	<b>1</b>
1.1 Background . . . . .	1
1.2 Motivation . . . . .	2
1.3 Thesis Objectives . . . . .	3
1.4 Thesis Outline . . . . .	3
<b>I State of the Art</b>	<b>5</b>
<b>2 Semantic Web Resources</b>	<b>6</b>
2.1 Generalities . . . . .	6
2.2 DBpedia . . . . .	7
2.3 GeoNames . . . . .	9
2.4 WordNet . . . . .	9
2.5 ImageNet . . . . .	10
<b>3 Disambiguation</b>	<b>11</b>
3.1 DBpedia Spotlight . . . . .	11
<b>4 Measures</b>	<b>12</b>
4.1 Distance measures . . . . .	12
4.2 Similarity measures . . . . .	12
<b>5 Existing services</b>	<b>13</b>
5.1 Web service 1 . . . . .	13
5.2 Web service 2 . . . . .	13

5.3	Annotation via stats . . . . .	14
<b>6</b>	<b>Conclusion</b>	<b>15</b>
6.1	Section 1 . . . . .	15
<b>II</b>	<b>Contribution</b>	<b>16</b>
<b>7</b>	<b>Proposed Methodology</b>	<b>17</b>
7.1	Section 1 . . . . .	17
<b>8</b>	<b>Proposed Architecture</b>	<b>18</b>
8.1	Technology Choices . . . . .	18
8.1.1	Java . . . . .	18
8.1.2	Neo4j . . . . .	18
8.1.3	NLP . . . . .	19
8.1.4	DBpedia Spotlight . . . . .	19
8.1.5	JAWS . . . . .	19
8.1.6	JENA . . . . .	19
8.1.7	JSoup . . . . .	20
8.2	Graph Structure . . . . .	20
8.2.1	Pro-Cons . . . . .	20
8.2.2	Vertexes . . . . .	20
8.2.3	Edges . . . . .	21
<b>9</b>	<b>Experiments</b>	<b>22</b>
9.1	Dataset . . . . .	22
9.2	Code Explanation . . . . .	22
9.3	Results and Analysis . . . . .	23
9.3.1	Evaluation methodology . . . . .	23
9.3.2	Graph-based experiments . . . . .	23
9.3.2.1	Direct Neighbors . . . . .	23
9.3.2.2	Lists - WL . . . . .	23
9.3.2.3	Lists - SL . . . . .	23
9.3.3	Plain-text experiments . . . . .	23
9.3.3.1	WikiLinks . . . . .	23
9.3.3.2	WikiContent . . . . .	23
<b>10</b>	<b>Conclusion</b>	<b>24</b>
10.1	Section 1 . . . . .	24
<b>A</b>	<b>Appendix Title Here</b>	<b>25</b>
	<b>Bibliography</b>	<b>26</b>

# List of Figures

# List of Tables



# Abbreviations

<b>SPARQL</b>	<b>SPARQL Protocol and RDF Query Language</b>
<b>RDF</b>	<b>Resource Description Framework</b>
<b>URI</b>	<b>Universal Resource Identifier</b>
<b>XML</b>	<b>eXtensible Markup Language</b>

# Chapter 1

## Introduction

### 1.1 Background

Image is a popular medium nowadays : it is easy to capture, can be really light on your computer and speaks to everyone without distinction of language.

In the all days life, people share their pictures on social networks in less than a blink of eye. In average, 70M of pictures are posted on Instagram each day and the users hit the “Like” button 2.5B times<sup>1</sup>. Other services like Picasa or Flickr exists but aren’t as used as Instagram which is the favorite in the eyes of the teen public.

Companies also produce a lot of media data. Industry companies need their products’ pictures, marketing and advertising studios use a lot of images in order to create new stuff for their client, . . . But the most consumer of media data are obviously mass media themselves : Newspapers, TV shows, news broadcasts are dealing with pictures at every moment of their day.

This huge production and consumption of images implies the need of an efficient way to store and search for the relevant one when the time comes. The best illustration to this need is to think of the nice but long moments one had with its relatives searching for the good picture of the new-born nephew in the family pictures album.

Since an image itself doesn’t have a natural plain-text representation the best way to

---

<sup>1</sup>Stats from : <https://instagram.com/press/>

describe it is to add meta-data (data about the data) such as its date of creation, its dimensions or, and this is what this thesis is about, some tags.

There are a lot of ways if one wants to annotate pictures. We can do it manually, using our own words (like “Dad”, “Home” . . . ), we can also analyze the raw picture, its pixel representation and compare some metrics (like the color histogram) to sample images in order to detect known concepts. Moreover, if the image already possesses annotations, we can enrich it semantically.

This field is so wide that it is impossible to speak about all the possibilities and technologies. In this study, we will focus on the last point and investigate the automation of the semantic enrichment. We will study the resources at our disposal and propose a solution keeping in mind the facts cited previously.

In the following section, we will present and discuss an application scenario to illustrate the motivation behind this thesis.

## **1.2 Motivation**

NewsTV is a famous TV news channel which runs 24/7 and only speaks about the current news. It has lot of reporters worldwide, covering the important local news and sending their production to the main site in Paris, France.

The employees often need to consult older coverages in order to explain the context of the news, to make the necrology of a famous actor who recently died or to re-use common shots. Therefor, they need to query the central multimedia database management system using keywords they are familiar with like “Elections, France, 2007, José Bové”. But sometimes, their research aren’t so specific and they are looking for more generic pictures, let say “Land, Tree, Animal”.

The first kind of keywords had been tagged by the former reporter who produced the coverage but he logically didn't think to add generic terms. NewsTV needs something to do it automatically when a picture, or any media, is first added to its system which a couple of initial tags.

Details about which kind of technology can be used to achieve this automatic tagging will come in the following sections. To summarize, the goal of this thesis is to propose a running prototype and evaluate different methods of tagging. The questions that we will try to answer during this study are described in the following section.

### 1.3 Thesis Objectives

Sed ullamcorper quam eu nisl interdum at interdum enim egestas. Aliquam placerat justo sed lectus lobortis ut porta nisl porttitor. Vestibulum mi dolor, lacinia molestie gravida at, tempus vitae ligula. Donec eget quam sapien, in viverra eros. Donec pel-lentesque justo a massa fringilla non vestibulum metus vestibulum. Vestibulum in orci quis felis tempor lacinia. Vivamus ornare ultrices facilisis. Ut hendrerit volutpat vulputate. Morbi condimentum venenatis augue, id porta ipsum vulputate in. Cur-abitur luctus tempus justo. Vestibulum risus lectus, adipiscing nec condimentum quis, condimentum nec nisl. Aliquam dictum sagittis velit sed iaculis. Morbi tristique augue sit amet nulla pulvinar id facilisis ligula mollis. Nam elit libero, tincidunt ut aliquam at, molestie in quam. Aenean rhoncus vehicula hendrerit.

### 1.4 Thesis Outline

The remainder of this thesis will be organized as follows :

**Chapter 2 - Semantic Web Resources:** presents different semantic web resources, their structures and how to browse them and how are they used in the literature.

**Chapter 3 - Disambiguation:** reviews the literature and assesses the most relevant ways to disambiguate a list of keywords which may be organize into sentences or not.

**Chapter 4 - Measures:** provides a solid background on semantic similarity and distance measures. We explore different metrics illustrating their pro/cons with examples.

**Chapter 5 - Existing Services:** describes existing image annotation services as well as their use-cases.

**Chapter 6 - State of the Art Conclusion:** summarizes the findings of the previous state of the art and opens the way to the presented contribution.

**Chapter 7 - Proposed Methodology :** presents the chosen methodology as well as some organizational points.

**Chapter 8 - Proposed Architecture :** details the technological choices by comparing them to their competitors and the chosen DBMS schema. Illustration figures will be presented.

**Chapter 9 - Proposed Architecture :** presents the chosen dataset, details some of the main algorithms and reviews the tests' results with the use of different evaluation methods.

**Chapter 10 - Contribution Conclusion :** summarize the findings of the presented research problem.

## **Part I**

# **State of the Art**

## Chapter 2

# Semantic Web Resources

Our study is focus on semantic enrichment of an initial set of keywords which can be organized as sentences or not. It is important to first understand what is a semantic concept and how concepts are organized into ontologies.

In this section, we will present some general notions about semantic concepts and review several semantic resources, their hierarchical structures and how we access them.

### 2.1 Generalities

Linguistic semantics is the study of meaning that is used for understanding human expression through language. It is easy for two human-being to communicate (given that they speak the same language) and to understand what their partner say even if he's using a tricky turn of phrase. However, this task becomes way more difficult when it comes to the comprehension of the human language by a machine. How can the computer guess that "I am totally dead" means in fact "I am really tired" and that the speaker isn't actually dead ? Machines need structured resources to understand us and the Semantic Web is one of them.

The notion of "Semantic Web" has been mentioned for the first time by Berners-Lee et al in [1]. In this paper, they describe it as a Web which is readable by machines in opposite

of most of Web's content which were designed for humans to read. The Semantic Web isn't a separate Web but an extension of the current one which will bring structure to the meaningful content of Web pages.

Two main technologies are used for the development of the Semantic Web : eXtensible Markup Language ( short XML) and the Resource Description Framework (short RDF). XML allows everyone to create their own tags and to arbitrary structure their documents but gives no information about what this structure means. Meaning is provided by RDF which stores it in sets of triples which are composed by a subject, a predicate and an object. Those three components can be related to the subject, the verb and object of an elementary sentence. In [2], Miller present a short introduction to the RDF standard and precise that a "Resource" can be any object which is uniquely identifiable by a Uniform Resource Identifier (URI).

The third basic component of the Semantic Web are collections of information called ontologies. An ontology is, in computer science, a document which defines the relations among concepts. Basically, Web ontologies are composed of a taxonomy, which defines classes of objects and their relations, and a set of inference rules.

In addition, the The New Oxford Dictionary of English defines the notion of "semantic concept" as : *An idea or thought that corresponds to some distinct entity or class of entities, or to its essential features, or determines the application of a term, and thus plays a part in the use of reason or language*

Given those basic notions, we will now further detail four semantic web resources, their taxonomies and review some of their usage found in the literature.

## 2.2 DBpedia

DBpedia is a project originally launched by two German universities (Berlin and Leipzig) and backed by an important community. It explore Wikipedia and extract



information from it which results on the creation of a multilingual, large-scale knowledge base. The extraction framework, all the available end-points as well as some facts and figures about the project are presented in [3].

DBpedia's ontology is based on classes (320 items) which form a subsumption hierarchy, the root element being `owl:Thing`, with a maximal depth of 5<sup>1</sup>. These classes are described by a total of 1650 different properties, forming a large set of RDF triples.

Even though DBpedia is now a worldwide project and provides pages in 125 languages, the English one is still the most represented. We can indeed find 4.58 million of things<sup>2</sup> including 1,445,000 instances of the class *Person*, 735,000 places *Place*, 251,000 *Species* ... The number of instances described in this language is about three time larger than the second and third language (French and German).

As well as any RDF-structured dataset, DBpedia can be requested with SPARQL (which is an recursive acronym : SPARQL Protocol and RDF Query Language) queries. SPARQL allows the user to search, add, modify or delete RDF data available on the Internet, see [4] for more details about the language.

```
1 PREFIX rdf: <http://www.w3.org/1999/02/22-rdf-syntax-ns#>
2 SELECT ?class
3 WHERE { <E> rdf:type ?class }
```

CODE 2.1: SPARQL Query : Search classes

```
1 PREFIX rdf: <http://www.w3.org/1999/02/22-rdf-syntax-ns#>
2 SELECT ?superClass
3 WHERE { <C> rdf:subClassOf ?superClass }
```

CODE 2.2: SPARQL Query : Search superclasses

Codes 2.1 and 2.2 present two simple and generic SPARQL search queries which return respectively the class(es) of a given entity E and the superclass(es) of a given class C.

---

<sup>1</sup>Complete classes tree : <http://mappings.dbpedia.org/server/ontology/classes/>

<sup>2</sup><http://wiki.dbpedia.org/about/facts-figures>

DBpedia also provides useful web services and HTTP endpoints. DBpedia Spotlight, which highlight DBpedia concepts in an input text is described in [5] and further details about disambiguation using this service are presented in section 3.1. The official DBpedia SPARQL endpoint<sup>3</sup> allows the user to send SPARQL query to the online Virtuoso Triple Store by using the browser interface or by sending a HTTP request. We learn in [3] that the average amount of hits per day of this endpoint is of 2,910,410 for the 3.8 dataset version.

## 2.3 GeoNames

GeoNames is a geographical database which contains more than 6,5 million places. It's a collaborative tool : users can add data or edit the existing ones. GeoNames' data are link to other RDF resources (mainly DBpedia).

Semantically speaking, each GeoNames' resource is link to an URI. This URI lead the user to an HTML page or to a RDF description. We can then write SPARQL queries using specific namespaces and request GeoNames. This resource is a key in our process, it will give us our first information about the picture given its GPS coordinates.

## 2.4 WordNet

WordNet is a lexical database created by Priceton University. Its structure is based on the "synset" (synonym set), a set of words with the same meaning. If DBpedia's architecture uses classes, Wordnet use hyponyms and hyperonyms. The root element is "entity".

This resource is useful if we are searching for entities (tsetse-fly, jaguar, ...) but it also contains more general concepts (vehicle, animal, ...). It also have interesting links, like meronyms ("HAS-PART" relation), holonyms (the opposite), synonyms or antonyms.

---

<sup>3</sup><http://dbpedia.org/sparql>

## 2.5 ImageNet

ImageNet is an image dataset based on the WordNet hierarchy. For each synset, ImageNet provide a set of images depicting it.

## **Chapter 3**

# **Disambiguation**

### **3.1 DBpedia Spotlight**

## Chapter 4

# Measures

### 4.1 Distance measures

Lorem ipsum dolor sit amet, consectetur adipiscing elit. Aliquam ultricies lacinia euismod. Nam tempus risus in dolor rhoncus in interdum enim tincidunt. Donec vel nunc neque. In condimentum ullamcorper quam non consequat. Fusce sagittis tempor feugiat. Fusce magna erat, molestie eu convallis ut, tempus sed arcu. Quisque molestie, ante a tincidunt ullamcorper, sapien enim dignissim lacus, in semper nibh erat lobortis purus. Integer dapibus ligula ac risus convallis pellentesque.

### 4.2 Similarity measures

Sed ullamcorper quam eu nisl interdum at interdum enim egestas. Aliquam placerat justo sed lectus lobortis ut porta nisl porttitor. Vestibulum mi dolor, lacinia molestie gravida at, tempus vitae ligula. Donec eget quam sapien, in viverra eros. Donec pellentesque justo a massa fringilla non vestibulum metus vestibulum. Vestibulum in orci quis felis tempor lacinia. Vivamus ornare ultrices facilisis. Ut hendrerit volutpat vulputate. Morbi condimentum venenatis augue, id porta ipsum vulputate in. Curabitur luctus tempus justo. Vestibulum risus lectus, adipiscing nec condimentum quis, condimentum nec nisl. Aliquam dictum sagittis velit sed iaculis. Morbi tristique augue sit amet nulla pulvinar id facilisis ligula mollis. Nam elit libero, tincidunt ut aliquam at, molestie in quam. Aenean rhoncus vehicula hendrerit.

## Chapter 5

# Existing services

### 5.1 Web service 1

Lorem ipsum dolor sit amet, consectetur adipiscing elit. Aliquam ultricies lacinia euismod. Nam tempus risus in dolor rhoncus in interdum enim tincidunt. Donec vel nunc neque. In condimentum ullamcorper quam non consequat. Fusce sagittis tempor feugiat. Fusce magna erat, molestie eu convallis ut, tempus sed arcu. Quisque molestie, ante a tincidunt ullamcorper, sapien enim dignissim lacus, in semper nibh erat lobortis purus. Integer dapibus ligula ac risus convallis pellentesque.

### 5.2 Web service 2

Sed ullamcorper quam eu nisl interdum at interdum enim egestas. Aliquam placerat justo sed lectus lobortis ut porta nisl porttitor. Vestibulum mi dolor, lacinia molestie gravida at, tempus vitae ligula. Donec eget quam sapien, in viverra eros. Donec pellentesque justo a massa fringilla non vestibulum metus vestibulum. Vestibulum in orci quis felis tempor lacinia. Vivamus ornare ultrices facilisis. Ut hendrerit volutpat vulputate. Morbi condimentum venenatis augue, id porta ipsum vulputate in. Curabitur luctus tempus justo. Vestibulum risus lectus, adipiscing nec condimentum quis, condimentum nec nisl. Aliquam dictum sagittis velit sed iaculis. Morbi tristique augue sit amet nulla pulvinar id facilisis ligula mollis. Nam elit libero, tincidunt ut aliquam at, molestie in quam. Aenean rhoncus vehicula hendrerit.

### 5.3 Annotation via stats

Sed ullamcorper quam eu nisl interdum at interdum enim egestas. Aliquam placerat justo sed lectus lobortis ut porta nisl porttitor. Vestibulum mi dolor, lacinia molestie gravida at, tempus vitae ligula. Donec eget quam sapien, in viverra eros. Donec pelentesque justo a massa fringilla non vestibulum metus vestibulum. Vestibulum in orci quis felis tempor lacinia. Vivamus ornare ultrices facilisis. Ut hendrerit volutpat vulputate. Morbi condimentum venenatis augue, id porta ipsum vulputate in. Curabitur luctus tempus justo. Vestibulum risus lectus, adipiscing nec condimentum quis, condimentum nec nisl. Aliquam dictum sagittis velit sed iaculis. Morbi tristique augue sit amet nulla pulvinar id facilisis ligula mollis. Nam elit libero, tincidunt ut aliquam at, molestie in quam. Aenean rhoncus vehicula hendrerit.

## Chapter 6

# Conclusion

### 6.1 Section 1

Lorem ipsum dolor sit amet, consectetur adipiscing elit. Aliquam ultricies lacinia euismod. Nam tempus risus in dolor rhoncus in interdum enim tincidunt. Donec vel nunc neque. In condimentum ullamcorper quam non consequat. Fusce sagittis tempor feugiat. Fusce magna erat, molestie eu convallis ut, tempus sed arcu. Quisque molestie, ante a tincidunt ullamcorper, sapien enim dignissim lacus, in semper nibh erat lobortis purus. Integer dapibus ligula ac risus convallis pellentesque.



## **Part II**

# **Contribution**

## Chapter 7

# Proposed Methodology

### 7.1 Section 1

Lorem ipsum dolor sit amet, consectetur adipiscing elit. Aliquam ultricies lacinia euismod. Nam tempus risus in dolor rhoncus in interdum enim tincidunt. Donec vel nunc neque. In condimentum ullamcorper quam non consequat. Fusce sagittis tempor feugiat. Fusce magna erat, molestie eu convallis ut, tempus sed arcu. Quisque molestie, ante a tincidunt ullamcorper, sapien enim dignissim lacus, in semper nibh erat lobortis purus. Integer dapibus ligula ac risus convallis pellentesque.

## Chapter 8

# Proposed Architecture

### 8.1 Technology Choices

#### 8.1.1 Java

Lorem ipsum dolor sit amet, consectetur adipiscing elit. Aliquam ultricies lacinia euismod. Nam tempus risus in dolor rhoncus in interdum enim tincidunt. Donec vel nunc neque. In condimentum ullamcorper quam non consequat. Fusce sagittis tempor feugiat. Fusce magna erat, molestie eu convallis ut, tempus sed arcu. Quisque molestie, ante a tincidunt ullamcorper, sapien enim dignissim lacus, in semper nibh erat lobortis purus. Integer dapibus ligula ac risus convallis pellentesque.

#### 8.1.2 Neo4j

Lorem ipsum dolor sit amet, consectetur adipiscing elit. Aliquam ultricies lacinia euismod. Nam tempus risus in dolor rhoncus in interdum enim tincidunt. Donec vel nunc neque. In condimentum ullamcorper quam non consequat. Fusce sagittis tempor feugiat. Fusce magna erat, molestie eu convallis ut, tempus sed arcu. Quisque molestie, ante a tincidunt ullamcorper, sapien enim dignissim lacus, in semper nibh erat lobortis purus. Integer dapibus ligula ac risus convallis pellentesque.

### 8.1.3 NLP

Lorem ipsum dolor sit amet, consectetur adipiscing elit. Aliquam ultricies lacinia euismod. Nam tempus risus in dolor rhoncus in interdum enim tincidunt. Donec vel nunc neque. In condimentum ullamcorper quam non consequat. Fusce sagittis tempor feugiat. Fusce magna erat, molestie eu convallis ut, tempus sed arcu. Quisque molestie, ante a tincidunt ullamcorper, sapien enim dignissim lacus, in semper nibh erat lobortis purus. Integer dapibus ligula ac risus convallis pellentesque.

### 8.1.4 DBpedia Spotlight

Lorem ipsum dolor sit amet, consectetur adipiscing elit. Aliquam ultricies lacinia euismod. Nam tempus risus in dolor rhoncus in interdum enim tincidunt. Donec vel nunc neque. In condimentum ullamcorper quam non consequat. Fusce sagittis tempor feugiat. Fusce magna erat, molestie eu convallis ut, tempus sed arcu. Quisque molestie, ante a tincidunt ullamcorper, sapien enim dignissim lacus, in semper nibh erat lobortis purus. Integer dapibus ligula ac risus convallis pellentesque.

### 8.1.5 JAWS

Lorem ipsum dolor sit amet, consectetur adipiscing elit. Aliquam ultricies lacinia euismod. Nam tempus risus in dolor rhoncus in interdum enim tincidunt. Donec vel nunc neque. In condimentum ullamcorper quam non consequat. Fusce sagittis tempor feugiat. Fusce magna erat, molestie eu convallis ut, tempus sed arcu. Quisque molestie, ante a tincidunt ullamcorper, sapien enim dignissim lacus, in semper nibh erat lobortis purus. Integer dapibus ligula ac risus convallis pellentesque.

### 8.1.6 JENA

Lorem ipsum dolor sit amet, consectetur adipiscing elit. Aliquam ultricies lacinia euismod. Nam tempus risus in dolor rhoncus in interdum enim tincidunt. Donec vel nunc neque. In condimentum ullamcorper quam non consequat. Fusce sagittis tempor feugiat. Fusce magna erat, molestie eu convallis ut, tempus sed arcu. Quisque molestie,

ante a tincidunt ullamcorper, sapien enim dignissim lacus, in semper nibh erat lobortis purus. Integer dapibus ligula ac risus convallis pellentesque.

### 8.1.7 JSoup

Lorem ipsum dolor sit amet, consectetur adipiscing elit. Aliquam ultricies lacinia euismod. Nam tempus risus in dolor rhoncus in interdum enim tincidunt. Donec vel nunc neque. In condimentum ullamcorper quam non consequat. Fusce sagittis tempor feugiat. Fusce magna erat, molestie eu convallis ut, tempus sed arcu. Quisque molestie, ante a tincidunt ullamcorper, sapien enim dignissim lacus, in semper nibh erat lobortis purus. Integer dapibus ligula ac risus convallis pellentesque.

## 8.2 Graph Structure

### 8.2.1 Pro-Cons

Lorem ipsum dolor sit amet, consectetur adipiscing elit. Aliquam ultricies lacinia euismod. Nam tempus risus in dolor rhoncus in interdum enim tincidunt. Donec vel nunc neque. In condimentum ullamcorper quam non consequat. Fusce sagittis tempor feugiat. Fusce magna erat, molestie eu convallis ut, tempus sed arcu. Quisque molestie, ante a tincidunt ullamcorper, sapien enim dignissim lacus, in semper nibh erat lobortis purus. Integer dapibus ligula ac risus convallis pellentesque.

### 8.2.2 Vertexes

Lorem ipsum dolor sit amet, consectetur adipiscing elit. Aliquam ultricies lacinia euismod. Nam tempus risus in dolor rhoncus in interdum enim tincidunt. Donec vel nunc neque. In condimentum ullamcorper quam non consequat. Fusce sagittis tempor feugiat. Fusce magna erat, molestie eu convallis ut, tempus sed arcu. Quisque molestie, ante a tincidunt ullamcorper, sapien enim dignissim lacus, in semper nibh erat lobortis purus. Integer dapibus ligula ac risus convallis pellentesque.

### 8.2.3 Edges

Lorem ipsum dolor sit amet, consectetur adipiscing elit. Aliquam ultricies lacinia euismod. Nam tempus risus in dolor rhoncus in interdum enim tincidunt. Donec vel nunc neque. In condimentum ullamcorper quam non consequat. Fusce sagittis tempor feugiat. Fusce magna erat, molestie eu convallis ut, tempus sed arcu. Quisque molestie, ante a tincidunt ullamcorper, sapien enim dignissim lacus, in semper nibh erat lobortis purus. Integer dapibus ligula ac risus convallis pellentesque.

## Chapter 9

# Experiments

### 9.1 Dataset

Lorem ipsum dolor sit amet, consectetur adipiscing elit. Aliquam ultricies lacinia euismod. Nam tempus risus in dolor rhoncus in interdum enim tincidunt. Donec vel nunc neque. In condimentum ullamcorper quam non consequat. Fusce sagittis tempor feugiat. Fusce magna erat, molestie eu convallis ut, tempus sed arcu. Quisque molestie, ante a tincidunt ullamcorper, sapien enim dignissim lacus, in semper nibh erat lobortis purus. Integer dapibus ligula ac risus convallis pellentesque.

### 9.2 Code Explanation

Sed ullamcorper quam eu nisl interdum at interdum enim egestas. Aliquam placerat justo sed lectus lobortis ut porta nisl porttitor. Vestibulum mi dolor, lacinia molestie gravida at, tempus vitae ligula. Donec eget quam sapien, in viverra eros. Donec pellentesque justo a massa fringilla non vestibulum metus vestibulum. Vestibulum in orci quis felis tempor lacinia. Vivamus ornare ultrices facilisis. Ut hendrerit volutpat vulputate. Morbi condimentum venenatis augue, id porta ipsum vulputate in. Curabitur luctus tempus justo. Vestibulum risus lectus, adipiscing nec condimentum quis, condimentum nec nisl. Aliquam dictum sagittis velit sed iaculis. Morbi tristique augue sit amet nulla pulvinar id facilisis ligula mollis. Nam elit libero, tincidunt ut aliquam at, molestie in quam. Aenean rhoncus vehicula hendrerit.

## **9.3 Results and Analysis**

### **9.3.1 Evaluation methodology**

Lorem ipsum dolor sit amet, consectetur adipiscing elit. Aliquam ultricies lacinia euismod. Nam tempus risus in dolor rhoncus in interdum enim tincidunt. Donec vel nunc neque. In condimentum ullamcorper quam non consequat. Fusce sagittis tempor feugiat. Fusce magna erat, molestie eu convallis ut, tempus sed arcu. Quisque molestie, ante a tincidunt ullamcorper, sapien enim dignissim lacus, in semper nibh erat lobortis purus. Integer dapibus ligula ac risus convallis pellentesque.

### **9.3.2 Graph-based experiments**

#### **9.3.2.1 Direct Neighbors**

#### **9.3.2.2 Lists - WL**

#### **9.3.2.3 Lists - SL**

### **9.3.3 Plain-text experiments**

#### **9.3.3.1 WikiLinks**

#### **9.3.3.2 WikiContent**



## Chapter 10

# Conclusion

### 10.1 Section 1

Lorem ipsum dolor sit amet, consectetur adipiscing elit. Aliquam ultricies lacinia euismod. Nam tempus risus in dolor rhoncus in interdum enim tincidunt. Donec vel nunc neque. In condimentum ullamcorper quam non consequat. Fusce sagittis tempor feugiat. Fusce magna erat, molestie eu convallis ut, tempus sed arcu. Quisque molestie, ante a tincidunt ullamcorper, sapien enim dignissim lacus, in semper nibh erat lobortis purus. Integer dapibus ligula ac risus convallis pellentesque.

## **Appendix A**

# **Appendix Title Here**

Write your Appendix content here.

# Bibliography

- [1] Tim Berners-Lee, James Hendler, Ora Lassila, et al. The semantic web. 2001.
- [2] Eric Miller. An introduction to the resource description framework. *Bulletin of the American Society for Information Science and Technology*, 25(1):15–19, 1998.
- [3] Jens Lehmann, Robert Isele, Max Jakob, Anja Jentzsch, Dimitris Kontokostas, Pablo N Mendes, Sebastian Hellmann, Mohamed Morsey, Patrick van Kleef, Sören Auer, et al. Dbpedia-a large-scale, multilingual knowledge base extracted from wikipedia. *Semantic Web Journal*, 5:1–29, 2014.
- [4] Eric Prud’Hommeaux, Andy Seaborne, et al. Sparql query language for rdf. *W3C recommendation*, 15, 2008.
- [5] Pablo N Mendes, Max Jakob, Andrés García-Silva, and Christian Bizer. Dbpedia spotlight: shedding light on the web of documents. In *Proceedings of the 7th International Conference on Semantic Systems*, pages 1–8. ACM, 2011.