MULTIMODAL SEMANTICS INTEGRATION USING ONTOLOGIES ENHANCED BY ONTOLOGY EXTRACTION AND CROSS MODALITY DISAMBIGUATION

AHMAD ADEL AHMAD ABU SHAREHA

UNIVERSITI SAINS MALAYSIA

2012

MULTIMODAL SEMANTICS INTEGRATION USING ONTOLOGIES ENHANCED BY ONTOLOGY EXTRACTION AND CROSS MODALITY DISAMBIGUATION

by

AHMAD ADEL AHMAD ABU SHAREHA

Thesis submitted in fulfillment of the requirements for the Degree of Doctor of Philosophy

MARCH 2012

ACKNOWLEDGEMENTS

IN THE NAME OF ALLAH THE ALL-COMPASSIONATE, ALL-MERCIFUL

All praises and thanks to Allah for giving me the energy and talent during my study. The favor, above all, before all, and after all, is entirely Allah's, to whom my never-ending thanks and praise are humbly due.

Gratitude beyond words to those who are always in my heart, my mother, my brother, my sisters and my uncle, who have always ushered me with love and support in every single step in my life. I will be grateful to them forever.

I would like to express my sincere appreciation to my supervisor Prof. Dr. Mandava Rajeswari for all the support and valuable guidance beyond the preparation of this thesis. I consider myself privileged to have had the opportunity to work under her guidance. I am also grateful to Dr. Dhanesh for all the assistance and the invaluable advices. I would also like to thank Dr. Latifur Khan of the University of Texas at Dallas for all the useful ideas he has given me from the discussions that we had. I would like to thank Yayasan Khazanah for sponsoring my study. Also, thanks for khazanah team, Dr. Ikmal, Ijlal, Shahrul, Fazlinda and Sariza whom been a tremendous source of support.

Last but not least, I would like to convey my appreciation to all my friends Mahmoud Jawarneh, Nor Idayu, Anush Achutan, Osama Alia, Alfian Abdulhalin, Hani Al-Mimi, Mohammad Mosa, Mahmoud Baklizi and Mozaherul Hoque for their help and support. Very special thanks to Mosleh Abu-Alhaj for the ultimate support.

.

LIST OF PUBLICATIONS

Journal Publications

Ahmad Adel Abu-Shareha, Rajeswari Mandava, Latifur Khan and Dhanesh Ramachandram ,“Multimodal Concept Fusion using Semantic Closeness for Image Concept Disambiguation” *Multimedia Tools and Applications (MTAP)*, Online first, 2011. doi:10.1007/s11042-010-0707-8.

Ahmad Adel Abu-Shareha, Mandava Rajeswari and Dhanesh Ramachandram, “Multimodal Integration (Image and Text) Using Ontology Alignment”, *American Journal of Applied Science*, 6(6), 2009, pp. 1217-1224.

Conference Publications

Ahmad Adel Abu-Shareha and Rajeswari Mandava, “Semantics Extraction in Visual Domain based on WordNet”, *Fifth FTRA International Conference on Multimedia and Ubiquitous Engineering*, Greece, Jun, 2011, pp.212-219.

Ahmad Adel Abu-Shareha, Rajeswari Mandava, Latifur Khan and Dhanesh Ramachandram, "Multimodal Concept Fusion using Semantic Closeness for Image Concept Disambiguation", **Best paper award,** *Advanced Future Multimedia Services (AFMS 2010)*, Korea, 2010.

Ahmad Adel Abu-Shareha, Mandava Rajeswari and Dhanesh Ramachandram, “SLADO-Semantic Lexical-based Alignment for Domain-specific Ontologies*”, International Technical Conference of IEEE Region 10 (TENCON)*, Singapore, 2009.

Ahmad Adel Abu-Shareha, Mandava Rajeswari and Dhanesh Ramachandram, "Two-way Dictionary-based Lexical Ontology Alignment", *International Conference on Computer Engineering and Application (ICCEA09)*, Manila, 2009, pp.151-157.

Ahmad Adel Abu-Shareha, Mandava Rajeswari and Dhanesh Ramachandram, “Image Concepts Disambiguation using associate Text Concepts”, *4th International Conference on Information Technology (ICIT 2009)*, Amman, 2009.

TABLE OF CONTENTS

CHAPTER 3- LITERATURE REVIEW

LIST OF TABLES

LIST OF FIGURES

LIST OF ABBREVIATIONS

CBIR	Content Based Image Retrieval
CIDOC-CRM	International Committee for Documentation-Conceptual Reference Model
CMRM	Cross Media Relevance Model
CRF	Conditional Random Field
CRM	Continuous Relevance Model
DM	Direct Matching
GA	Genetic Algorithm
IAPR TC-12	International Association of Pattern Recognition, Technical Committee 12
LBO	Level-Based Overlapping
MBRM	Multiple Bernoulli Relevance Model
M-OWL	Multimedia- Web Ontology Language
MPEG-7	Moving Picture Experts Group ISO/IEC standard
MSI	Multimodal Semantics Integration
NLP	Natural Language Processing
OWL	Web Ontology Language
POS	Part Of Speech
RwD	Retrieval-with-Diversity
UMLS	Unified Medical Language System
WNS	WordNet Similarity
WSD	Word Sense Disambiguation

INTEGRASI SEMANTIK MULTIMODAL MENGGUNAKAN ONTOLOGI-ONTOLOGI DIPERTINGKATKAN MELALUI PENGEKSTRAKAN ONTOLOGI DAN PENYAH-KEPENDUAAN PERSILANGAN MODALITI

ABSTRAK

Peningkatan jumlah data multimodal seperti dokumen tekstual, imej beranotasi dan halaman sesawang telah mewujudkan keperluan bagi teknik pemanipulasian bagi data-data tersebut. Kelemahan ciri-ciri peringkat rendah imej dan teks adalah satu isu utama kerana lazimnya, ciri-ciri ini tidak mencukupi untuk pemanipulasian data. Oleh itu, memperoleh maklumat mencukupi dan bererti dari data multimodal, dan kemudiannya menggunakan maklumat berkenaan secara sesuai amat penting bagi pemanipulasian data. Tesis ini mencadangkan suatu proses integrasi semantik multimodal (MSI) bagi mengekstrak dan menggabung semantik dari modaliti tekstual dan imej, dan kemudian menggunakan gabungan ini bagi pemanipulasian data. Proses yang dicadangkan pertamanya mengekstrak perwakilan tekstual dari modaliti tekstual dan imej, diikuti pemetaan perwakilan ini kepada beberapa konsep dalam suatu sumber pengetahuan yang lebih kaya menggunakan sub-proses penjajaran berasaskan semantik. Penyah-kemenduaan persilangan modaliti kemudian dijalankan menggunakan keterhampiran semantik bagi memperoleh suatu set semantik tertambah baik. Akhir sekali, set semantik ini digabungkan bagi menghasilkan maklumat yang kaya dan lengkap berdasarkan sumber-sumber tergabung. MSI telah dinilai ke atas dua tugas, iaitu penyah-kemenduaan dan dapatan semula dengan kepelbagaian (RwD), menggunakan 20,000 contoh multimodal dari set data ImageCLEF. Dalam penilaian pertama, MSI berjaya meningkatkan kepersisan input-input berkependuaan sebanyak 32% berbanding kaedah konvensional, sementara mengekalkan kadar panggil balik. Bagi RwD pula, kepelbagaian penyelesaian yang diperolehi telah dipertingkatkan sebanyak 12% sementara mengekalkan ketepatan. Kaedah bukan-berasaskan-kepelbagaian juga telah meningkatkan kepersisan dapatan semula berbanding kaedah-kaedah sedia ada. Hasil eksperimen menunjukkan bahawa setiap komponen MSI telah mewajarkan pilihan untuk membina dan menggunakan komponen-komponen yang dipilih di dalam proses keseluruhan.

MULTIMODAL SEMANTICS INTEGRATION USING ONTOLOGIES ENHANCED BY ONTOLOGY EXTRACTION AND CROSS MODALITY DISAMBIGUATION

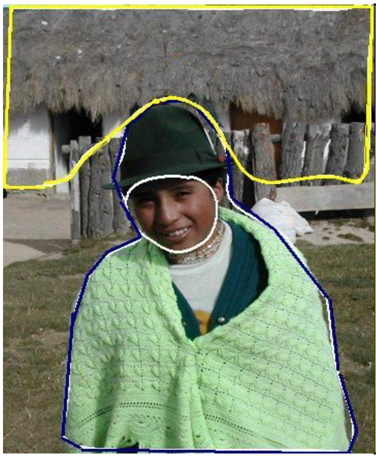
ABSTRACT

The increasing amount of multimodal data such as text documents, annotated images and web pages have necessitated the development of effective techniques for their manipulation. The ineffectiveness of low-level image and textual features is one of the main issues as these features are commonly insufficient for effective data manipulation. Therefore, obtaining sufficient and significant information from the multimodal data, and then further using this information in the proper manner is penultimate in data manipulation tasks. This thesis proposes a multimodal semantics integration (MSI) process to extract and integrate the semantics from the image and text modalities, and to use these semantics for manipulation tasks. The proposed process firstly extracts a textual representation from the textual and image modalities, followed by mapping the representation to concepts in a condensed knowledge source using a semantic-based alignment sub-process. Cross modality disambiguation is then performed using semantic closeness to obtain a set of enhanced semantics. Finally, the extracted and enhanced semantics are combined to deliver rich and sufficient information based on the integrated sources. MSI was evaluated on two tasks, namely disambiguation and retrieval-with-diversity (RwD), using 20,000 multimodal instances from the ImageCLEF dataset. In the disambiguation task, MSI improved the precision of ambiguous inputs by 32% over the conventional approach while preserving recall. In the RwD task, the diversity of the obtained solution was improved by 12% over the non-diversity-based approach while maintaining accuracy. The proposed non-diversity-based approach also improved the precision of the retrieval task by over the state-of-the-art approaches. Experimental results further showed that each proposed component of MSI justified the choice for building and utilizing the selected components within the overall process.

CHAPTER ONE

INTRODUCTION

Multimodal data is the form of data that combines multiple modalities in a single entity. In the content of this thesis, multimodal refers to the data that consist of text passage(s) and image(s) (Jiang and Tan, 2009). Examples of multimodal (image-text) data are scientific documents, annotated images, and web pages. Figure 1.1 gives an example of such data. Multimodal data have been utilized in a variety of communication models, mostly in education, medicine, advertising and industry. This growth in multimodal data has been supported by advances in telecommunication technologies, the Internet, computational power, and storage capacity. Recent preoccupation with utilizing multimodal data has made this form of data attractive, widespread, and broadly shared. In light of the above, there has been a dire need to manipulate such data (Christel et al., 1998; Jaimes et al., 2005; Stewart and Kowaltzke, 2007).



a dark-skinned, dark-haired girl wearing a black hat, a grey tee-shirt, a dark green vest and a light green poncho, is standing in the backyard of a house; a white painted house with a thatched roof and a wooden fence in the background; there is also a plastic bag in front of the fence.

Figure 1.1: Examples of multimodal (image-text) data (Grubinger et al., 2006)

1.1 Background: Multimodal Data Manipulation

Automatic multimodal data manipulation in web pages, personal images, medical cases, and library records is in great demand because it reduces human labor in managing huge data archives.

Generally, data mining and manipulation, such as classifying, clustering, indexing, and retrieval, are based on the *information* extracted from the data. Indeed, owing to their richness and the diversity of their representation, manipulating multimodal data is challenging because images are represented by pixels whereas the texts are represented by words and phrases. More importantly, multimodal data manipulation over the Internet is quite challenging because the amount of data being shared is growing on a day-to-day basis (Atika et al., 2009).

Applications of multimodal manipulation, such as the *web search engine*, the most popular multimodal data manipulation tool on the Internet, typically process text only. Similarly, common approaches for annotated image retrieval (Wilkins et al., 2005) and medical image retrieval (Costa et al., 2009) depend solely on the text part. Basically, the solely text-based approach, which employs words only, has been established to ease the manipulation of multimodal data. In fact, text-based manipulation and textual query formulation are much simpler and faster than using visual images. The solely text-based approach achieves the desired goal and yields sufficient output (Luo et al., 2003).

With the acceleration and increase in the shared information over the Internet, the solely text-based approach has encountered an emerging information overload problem. Information overload is the inability of a machine to make a correct decision, especially with regards to retrieval over the Internet, due to the presence of a huge amount of information and the utilization of low informative presentations of words only (Montebello, 1998). As such, the *efficiency of the aforementioned solely text-based approach has diminished* (Middleton and Baeza-Yates, 2007).

Basically, the efficiency of the manipulation approaches can be increased in two ways: the first depends on increasing the sources of information, whereas the second is based on enhancing the quality of the extracted information.

In *increasing the sources of information*, alternative approaches to multimodal manipulation reported in the literature utilize embodied image features, with the text words, to enhance the performance of the manipulation applications. For this purpose, feature concatenation and data fusion have been utilized to efficiently combine the extracted information. Unfortunately, this concatenation, although richer than the solely word-based approach, does not solve the problem completely because it continues to depend on low-information features and words (Zhao and Grosky, 2002; Kuo et al., 2005; Lacoste et al., 2007).

In *enhancing the quality of the extracted information*, the image features and words in text have to be replaced with more valuable information. Information varies in terms of style and informative ability. In particular, information in a low-level form, such as image features and words in text, is extracted directly from the data. However, high-level information, such as object identities in the image and vocabulary, is extracted by interpreting the low-level features using prestored associations of low-level and high-level information.

Clearly, information at the high level is more informative; however, it is more complicated and challenging because it requires a complex transformation process and a suitable prestored association.

Multimodal Semantics

Semantics is a high-level form of information that mimics the human model in describing the content of the data. Semantic-based applications interpret the machine-extracted low-level features using prestored knowledge. Basically, semantics are extracted by transferring the features into components in the utilized knowledge source. Fortunately, although with enormous challenges, text processing (Dietze and Domingue, 2009) and image processing (Wang et al., 2005) approaches and applications have evolved into semantics by using embryonic low-level features. This semantic revolution has been supported by the availability of knowledge resources in different fields (Sheth, 1995; Amato and Lecce, 2008). Unfortunately, semantics extraction from multimodal data faces huge obstacles related to challenges of semantics extraction from its underlying modalities; these challenges are summarized below:

First, semantics extractions in both image and text are ambiguous and not firm. In an image, this ambiguity is due to the fact that the low-level features that can be extracted directly from the image are mostly not discriminative. As a matter of fact, many objects and scenes share the same low-level features; therefore, mapping features into semantics is surrounded with huge issues of ambiguity. In text, vocabulary mismatching, such as when the same meaning can be expressed using different words and several meanings can be expressed using the same set of words, makes the extracted semantics of the text ambiguous as well. As such, in multimodal semantics extraction, combining ambiguous sources of information produces poor results.

Second, limitations related to the available knowledge sources utilized with the semantics extraction process exist in both image and text. Such limitations are embodied in the strength aspect because most of the available knowledge sources are upper-level and contain general and unfocused knowledge. The unfocused nature of the knowledge is characterized by giving all the possible interpretations of a given item. This type of knowledge does not provide precise semantics to the data being interpreted. This being the case, multimodal semantics that can be extracted based on such knowledge sources are not highly informative and not efficient for use in the manipulation tasks.

Third, no common knowledge source exists for the diversely represented image and text can combine features with words.

In summary, information extraction from the underlying modalities in multimodal data faces challenges in acquiring sufficient and efficient information in the form of semantics. In essence, semantic extraction from multimodal data does not have a suitable foundation to establish a good outcome. Consequently, the multimodal approaches continue to depend mainly on low-level features for manipulation tasks.

1.2 Problem Statement and Research Questions

The problems of how to extract sufficient and richer information from multimodal data and how to use the extracted information in the manipulation tasks, as mentioned earlier, remain unsolved. Overcoming such obstacles can be studied from different perspectives. Indeed, the problem statement of the present thesis is formulated around bridging this gap:

- How to *extract* unified and sufficient semantics content from the image and the text in multimodal data, and how to utilize the unified semantics in the multimodal manipulation tasks.

Consequently, the problem of extracting and utilizing such semantics can be divided into several subproblems:

- How to *transfer* both image and text from the machine-extracted features into the semantics of an identical representation.
- What *knowledge source* can be utilized for the semantic extraction from the image and text.
- How to *disambiguate* and enhance the extracted semantics from the image and the text in multimodal data.
- How to *combine* the enhanced semantics of the image and the text.
- How to *use* the unified extracted semantics for manipulation purposes.

1.3 Goal and Objectives

The main goal of this research is to propose a multimodal semantics integration process that can extract the semantic content of the underlying modalities, utilize the semantics of each modality to disambiguate the semantics of the other, and, finally, combine the semantics of the underlying modalities in a unified output.

As such, the extracted semantics of the multimodal data sufficiently represent the content of both modalities based on a utilized conceptual knowledge. For example, the integrated semantics output of the example in Figure 1.1 is illustrated in Figure 1.2. In Figure 1.2 the concepts are represented in boxes, whereas the arrows represent the relationships.

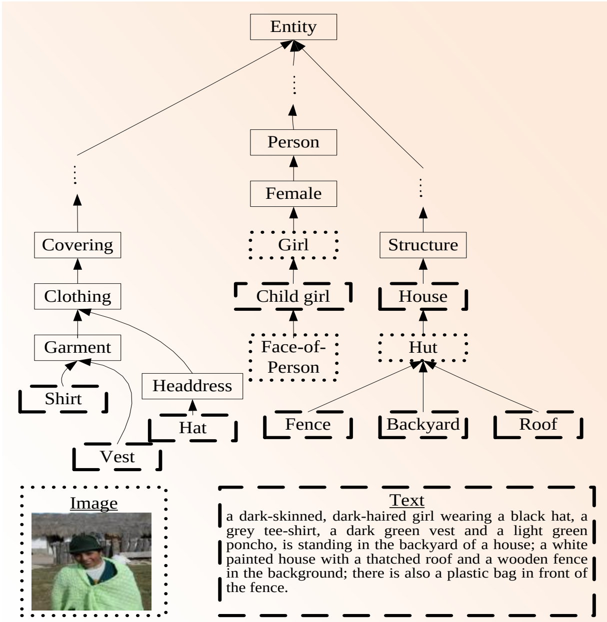


Figure 1.2: Example of multimodal semantics

The objectives of the research are listed below:

- To *construct* a domain ontology with a focused semantics to overcome the generality of the available knowledge sources.
- To *transfer* both image and text machine-extracted information into semantics based on a given ontology.
- To *disambiguate* and improve the semantics of the image and text modalities in multimodal data.
- To *integrate and combine* the semantics of the image and the text modalities in multimodal data.
- To employ the semantics of a multimodal data in a *retrieval task.*

1.4 Methodology

Generally, given a multimodal instance, the proposed *multimodal semantics integration* transfers each of the underlying modalities independently to an equivalent form of semantics. The semantics of each modality are then used to ease the ambiguity and shortcomings of the other modality. Then, based on their semantics, the modalities are integrated over the domain knowledge. The utilized domain knowledge is created by identifying the common semantics for the domain elements. The major processing steps of the proposed *multimodal semantics integration*, as illustrated in Figure 1.3, are:

- Domain ontology extraction: construct a domain ontology with a focused semantics.
- Semantics extraction: extract a semantics from both of image and text based on the extracted ontology.
- Cross modality disambiguation: disambiguate each modality based on the other.
- Integration: combine the disambiguated modalities.
- Utilization: use the integrated semantics in manipulation tasks.

Domain ontology extraction is implemented independently from the actual multimodal data integration processes. The domain ontology extraction process identifies the common semantics of the domain elements and then chooses the semantics of each element subsequently. As the domain ontology is extracted, the semantic-based processes over the image and text modalities, which include extraction, disambiguation and combination, are executed based on the extracted ontology.

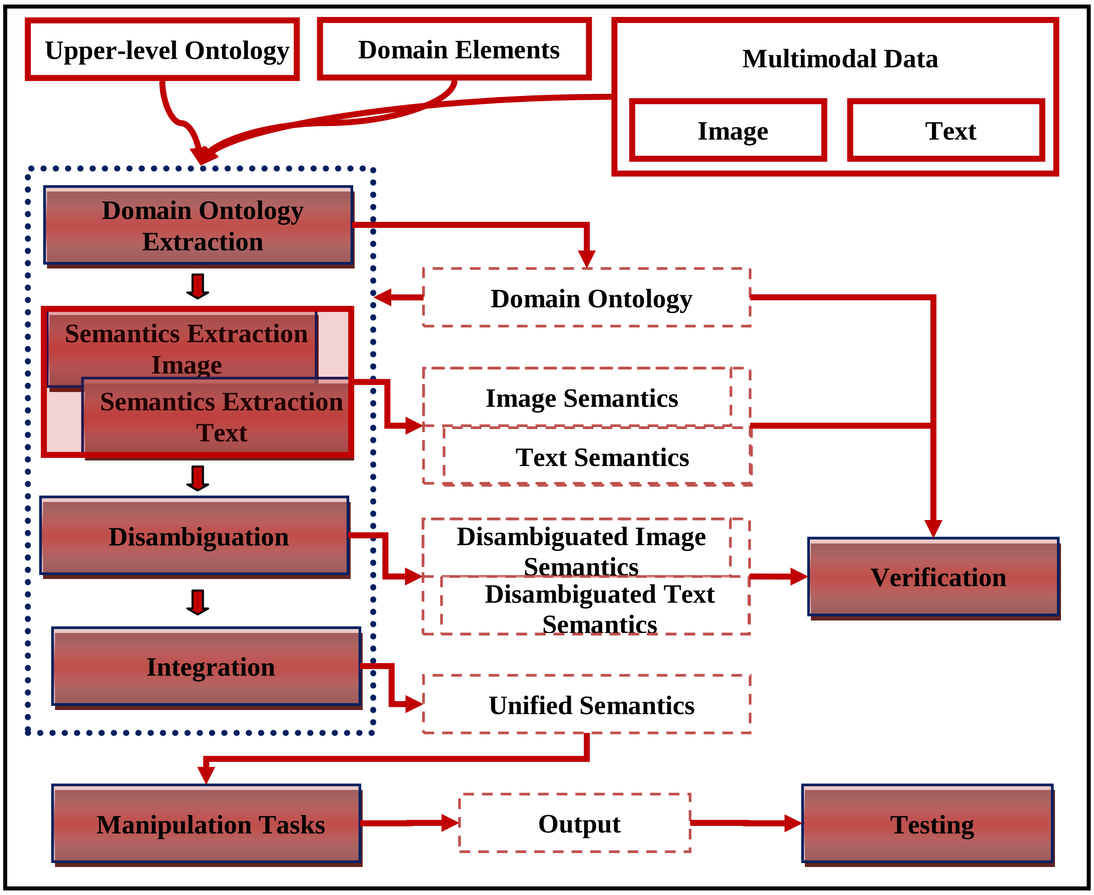


Figure 1.3: Research Methodology

The image and text machine-extracted information are transformed into semantics. Then, the text semantics is used, in the disambiguation process, to disambiguate the image semantics and vice versa. Finally, the output is constructed by combining the semantics of both modalities. Generally, the proposed *multimodal semantics integration* is designed based on a set of previously described processes. The set of processes are utilized collaboratively to achieve the overall goal. However, each process can be utilized independently to achieve a specific task in a given application. Therefore, each process is designed, created, tested individually and independently and verified before it is integrated in the process. Thus, the overall process is built and implemented by adding a single component at a time.

1.5 Scope

The scope of the current research is limited to proving the concepts of the proposed *multimodal semantics integration* and the other concepts highlighted in the objectives. To be more specific, topics that might overlap with the proposed work are not covered. The following points should thus be pointed out:

- Mainly, the proposed work does not cover image and text data that have no correlations with each other.
- The proposed work only covers noise-free multimodal data on natural images and grammatically error-free text; noise and clutters are not covered. The proposed work also does not cover preprocessing stages of data scrubbing that have no effect on the overall outcome.
- The proposed work does not dwell on text and image processing issues.
- Image content extraction and disambiguation are given more weight in the current thesis. For one thing, the aim of the current work is to show the potential of the image modalities in the multimodal data. For another, the purpose is to show a different perspective of multimodal data because most state-of-the-art approaches dwell on the textual information.
- The negative part of the text is totally ignored, and the process assumes that all the provided information is positive.

Actually, given the goal and the scope of the current research, the available dataset that could be used for testing and verification is very limited. Fortunately, the ImageCLEF dataset fits in the domain of this research because it provides a set of annotated images consisting of still images and free texts (Grubinger et al., 2006).

1.6 Impact of the Study

The impact of the multimodal semantics integration process is the ability to fit, after adding data-specific preprocessing, in any multimodal data, such as medical cases, scientific documents, and annotated images, given upper-level knowledge, such as WordNet or domain-specific ontologies. The output can be further processed to fit the tasks on hand, including image–text relationships, retrieval, Question-Answering (QA) and others.

1.7 Contributions

The main contributions of this thesis are as follows:

1. Presenting and experimenting with a new approach for *multimodal semantics integration* at the concept-level, based on semantic closeness.
2. Establishing a new approach for semantics-based lexical *alignment* that transfers the image and text machine-extracted features into semantic concepts in the utilized knowledge source.
3. Providing a new approach for *domain knowledge building* over WordNet.
4. Establishing a new approach for cross-modality disambiguation using semantic closeness.
5. Setting a new approach for annotated image *retrieval-with-diversity* based on multimodal semantics integration.

1.8 Organization of the Thesis

This thesis is organized into ten chapters as follows: **Chapter One** has introduced the characteristics, significance, and challenges of information extraction from multimodal data and the desire for extraction semantics from such data. Also, this chapter has provided insight into the research problems to be addressed throughout the thesis. The goals and objects, methodology, contributions, and scope of the proposed multimodal semantics integration have also been introduced.

Chapter Two introduces the fundamental concepts of the semantic extraction process, together with a description of their procedures. Furthermore, the use of ontologies as forms of knowledge in the semantics extraction process is discussed. **Chapter Three** discusses, in general, state-of-the-art semantics extraction from image, text, and the state-of-the-art multimodal manipulation. **Chapter Four** explains the proposed multimodal semantic integration process, its characteristics and significance, elements, inputs, and forms of output. **Chapter Five** presents the proposed mechanisms that extract textual features from image and text as the first step in the overall multimodal semantics integration process. **Chapter Six** presents the proposed alignment process as the components that transfer the machine-extracted features into semantic concepts. The alignment application in the field of ontology alignment is presented, and the output results are highlighted. **Chapter Seven** presents the proposed method responsible for automatically extracting the domain ontology. Chapter Seven presents the implementation of the domain extraction method over few datasets. **Chapter Eight** presents the proposed semantic closeness method which carries out the disambiguation processes and the obtained results. **Chapter Nine** presents the output and the experiments conducted over the overall proposed process. Finally, **Chapter ten** offers conclusions and directions for future research.

CHAPTER TWO

BACKGROUND ON SEMANTICS EXTRACTION

This chapter discusses the theoretical background of the semantics extraction process. After the introduction, the notions of knowledge sources, ontologies and WordNet are given subsequently. The semantics extraction procedures are described next. Then, the use of ontologies and WordNet in the semantics extraction process is discussed. A conclusion is finally provided at the end of this chapter.

2.1 Introduction

For humans, semantics denotes what is acquired by interpreting visual or verbal inputs based on previous knowledge. For the machine, semantics is acquired by interpreting data (e.g., image, text, database) in a standard form (Nielson and Nielson, 1992; Obitko et al., 2010). The standard form in which the semantics is presented consists of predefined tags and relationships that are inclusively stored in a *knowledge source*.

2.2 Knowledge Sources

A *knowledge source* is identified when "*any intelligent entity that wishes to reason about its world encounters an important, inescapable fact*" (Davis et al., 1993). In a knowledge source, the body of knowledge consists of a set of facts (i.e.: tags and relationships) that are stored in a knowledge base (Guarino and Giaretta, 1995). These facts are represented in a standard form using one of the knowledge representation schemes that makes such knowledge useable.

There are several *knowledge representation schemes* that have been proposed in the field of knowledge engineering, such as logical representation (Baral and Gelfond, 1994; Davis, 1993), productive rules (Vickery, 1993) and semantic networks (Steyvers and Tenenbaum, 2005). Each representation scheme has its own syntax and semantics. The syntax of the representation scheme is embodied in a list of predefined tags and relationships which allows knowledge engineers to encode knowledge. The semantics of the representation scheme is inferred from the definitions and the meaning of the defined tags and relationships, thus allowing knowledge to be inferred and utilized. An example of a representation scheme in First Order Logic (FOL) (Davis, 1993) is illustrated in Figure 2.1.

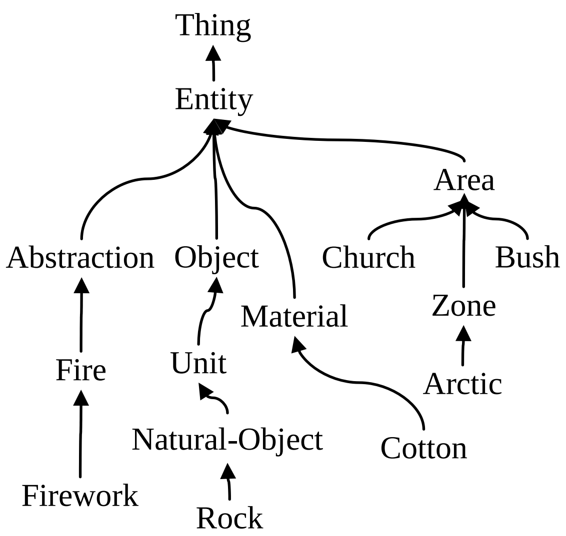
Tags	**Relationships**
M(x) for x is a Male C(x) for x has a child F(x) for x is a Father	$\forall$ (x)(M(x) ^C(x)$\rightarrow$ F(x))

Figure 2.1: Knowledge representation in First Order Logic

2.3 Ontologies

Ontology is a *knowledge representation* that is founded based on the notion of concepts. A *concept* is a tag that is identified by a word, a phrase or a label. Generally, there is no specific and widely accepted definition of what ontology is. However, there are two conditions that should be satisfied in order to call a knowledge source an ontology. First is the *conceptualization principle*, which means that the domain elements should be described by concepts (e.g., real names or abstract ideas). The second is the *categorizing principle*, by which the domain concepts are categorized using hierarchical relationships. The hierarchical relationships connect a general concept e.g. “Material” with its specific concept e.g.

"Cotton" and vice versa. Figure 2.2(a) illustrates an example of an ontology. The strings are concepts and the arrows are the relationships. Figure 2.2(b) illustrates the logical form of the ontology represented by the graphical form in Figure 2.2(a). The logical form can be built using one of the representation schemes mentioned earlier (Meersman, 2001; Waterson and Preece, 1999).




```
<?xml version="1.0"?>
<rdf:RDF
   xmlns:rdf="http://www.w3.org/1999/02/22-rdf-syntax-ns#"
   xmlns:xsd="http://www.w3.org/2001/XMLSchema#"
   xmlns:rdfs="http://www.w3.org/2000/01/rdf-schema#"
   xmlns:owl="http://www.w3.org/2002/07/owl#"
   xmlns="http://www.owl-ontologies.com/unnamed.owl#"
 xml:base="http://www.owl-ontologies.com/unnamed.owl">
 <owl:Ontology rdf:about=""/>
 <owl:Class rdf:ID="Unit">
   <rdfs:subClassOf>
     <owl:Class rdf:ID="Object"/>
   </rdfs:subClassOf>
 </owl:Class>
 <owl:Class rdf:ID="Bush">
   <rdfs:subClassOf>
     <owl:Class rdf:ID="Area"/>
   </rdfs:subClassOf>
 </owl:Class>
 <owl:Class rdf:ID="Natural-Object">
   <rdfs:subClassOf rdf:resource="#Unit"/>
 </owl:Class>
```

(a)The graphical-form of an ontology **(b)The logical-form of an ontology**

Figure 2.2: Examples of ontology

Ontologies are used in major fields such as Artificial Intelligence, Semantic Web, Software Engineering, Biomedical Informatics and Information Architecture (W3C, 2004 ; Ontology Works Inc, 2007). The *advantages* of using ontology as a knowledge representation on the push side are the capability of giving a standardization characteristic for the knowledge being represented and enhancing the data quality through predefined semantics, tags and reasoning support. On the pull side, it is more flexible for the automated construction and enrichment using any informative source such as the internet (Meersman, 2001).

There are two types of ontologies, the *domain-specific ontology* and *upper level ontology*. *Domain-specific ontology* encodes reusable *domain* concepts and represents the semantics of an established *domain*. *Upper level* ontology encodes the concepts from diverse and general domains, and covers the semantics of wide and undetermined domains. Because it is hard to determine the boundary of a specific domain and which concepts should be included in its ontology and which should not, consequently, most of the existing ontologies are an upper-level ontology, which covers a general and wide domain. One of the most utilized upper level ontologies is WordNet, which is described below.

2.4 WordNet

WordNet is a lexical resource, dictionary, thesaurus and knowledge source for the English language. As a dictionary, WordNet defines and briefly describes the words in the English language. As a knowledge source, WordNet has been built based on a set of synsets. The synset is a group of words with similar meanings and connected with each other by several relationships such as hypernymy and hyponymy. The hypernymy relationships connect a synset to its associated synsets with a specific granularity, such as connecting the concept “animal” to the concepts “mouse”, “cat” and “horse”. In contrast, the hyponymy relationship connects specific synsets to their general ones.

WordNet is considered as an ontology based on its noun synsets, hypernymy and hyponymy relationships. The *conceptualization and categorization conditions* of the ontology are satisfied in WordNet as follows: the *nouns* themselves satisfy the *conceptualization principle* of the ontology, while the *hierarchical structure*, which is formed by the hypernymy and hyponymy relationships, satisfies the *categorization principles*. Figure 2.3 illustrates a part of the noun hierarchy in WordNet (Fellbaum, 1998). The *advantages* of using WordNet are embodied in encapsulating all the English words and grouping the words of similar meaning in synsets.

8 senses of wood
Sense 1
wood -- (the hard fibrous lignified substance under the bark of trees)
=> plant material -- (material derived from plants)
=> material, stuff -- (the tangible substance that goes into the makeup of a physical object; "coal is a hard black material"; "wheat is the stuff they use to make bread")
=> substance, matter -- (that which has mass and occupies space; "an atom is the smallest indivisible unit of matter")
=> physical entity -- (an entity that has physical existence)
=> entity -- (that which is perceived or known or inferred to have its own distinct existence (living or nonliving))

Sense 2
forest, wood, woods -- (the trees and other plants in a large densely wooded area)
=> vegetation, flora, botany -- (all the plant life in a particular region or period; "Pleistocene vegetation"; "the flora of southern California"; "the botany of China")
=> collection, aggregation, accumulation, assemblage -- (several things grouped together or considered as a whole)
⋮

Figure 2.3: Part of the noun hierarchy in WordNet for the word "wood"

2.5 Semantics Extraction

Semantics extraction is a process of deducing semantics (e.g., tags, relationships or semantics relatedness) from a given input data with reference to an associated knowledge source. This process involves two steps: mapping and mining, as illustrated in Figure 2.4. The mapping procedure maps the input data to entities in the given knowledge source, while the mining procedure identifies the final semantics output from the knowledge source, given the mapped entities.

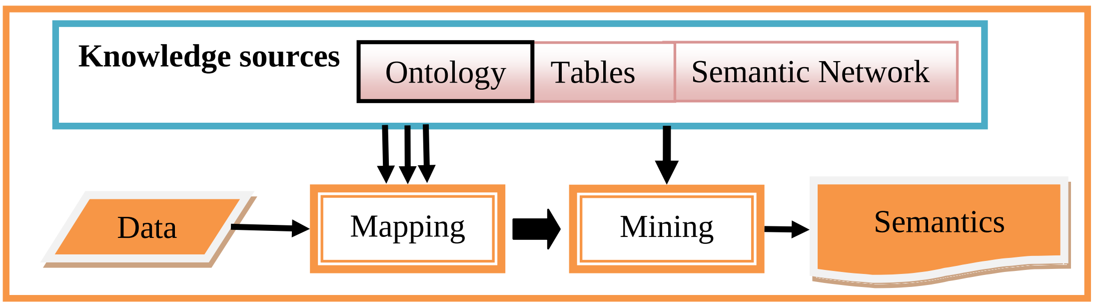


Figure 2.4: The semantics extraction process

Example

Figure 2.5 illustrates an example of an ideal and non-typical process of semantics extraction from image data, which is described as follows:

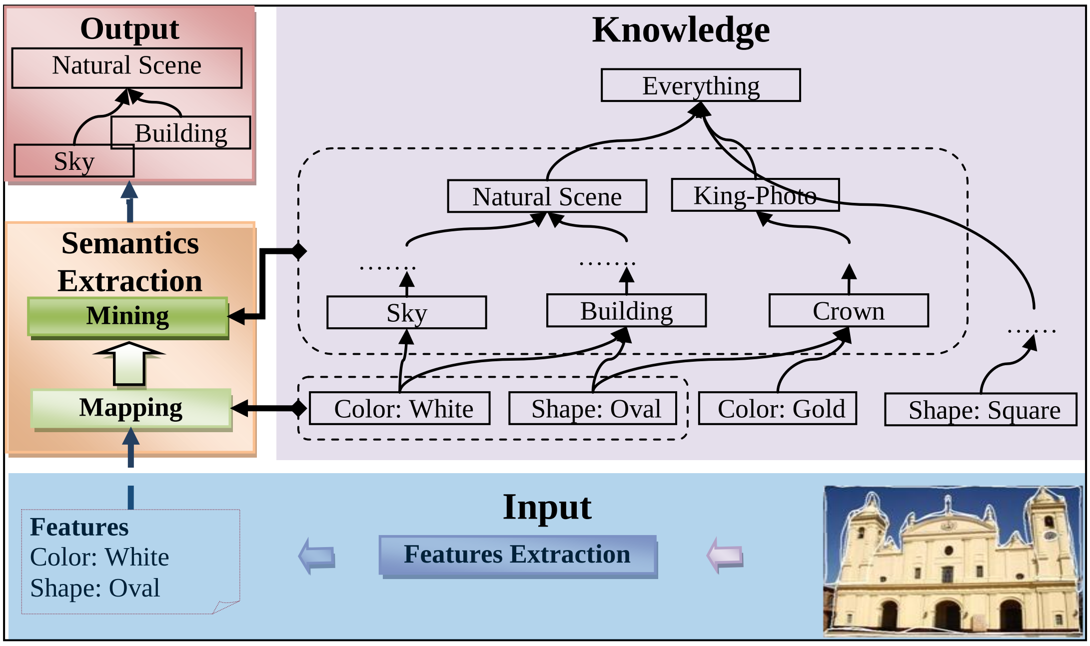


Figure 2.5: An example of the semantics extraction process from image data

First, the input image features are projected over the knowledge source. Second, the input elements are mapped into entities in the knowledge source. As such, the entities in the knowledge source which have equal values to the inputs features in the example, white and oval, are identified using word mapping procedure. Third, the mining procedure is executed over the identified entities and extracts the output.

In the example, the mining procedure is a rule-based flooding, which transfers from one level to another in the knowledge source by tracking the relationships from the identified entities up to the root node (e.g., everything). The final semantics output in this example is extracted subsequently as the parent concept (other than the root) that is connected to the maximum number of concepts in the track. As such, in the example above, the output is “natural scene”.

2.5.1 Mapping Procedure

Mapping procedure depends on the form of the input data. If the inputs are numerical values, mapping is carried through using mathematical operators such as "equal", "greater than" or "less than". If the inputs are words, a string matching method is utilized.

In the literature, semantics extraction from textual data has used a direct mapping procedure. The direct mapping of the textual data is facilitated by supplying a knowledge source that can fit adequately with the expected inputs (Varelas et al., 2005). With an image, if the expected inputs are limited, a direct mapping is also used (Jin et al., 2010), while, a classification technique is used if the range of the expected inputs is wide (Penta et al., 2007).

The mapping procedure, despite its form, can be carried through only if the data can be compared and matched with the entities in the associated knowledge source. To ensure that the mapping can be executed, the knowledge and the data should be harmonized. The overall harmony of the data and knowledge is determined in three elements: coverage, representation and granularity.

The *representatio*n is the form of the data, such as symbols, numbers and words. The representation of the knowledge entities and the data should be identical in order to allow the mapping procedure to be executed.

To fulfill the harmony in *coverage*, the domain of the utilized knowledge source has to cover all the possible values that the data may have. Generally, the first step to ensure harmony in coverage is to identify all the possible data values, then to find or build a suitable knowledge source to encapsulate those values.

Granularity is the level of detail at which the data is presented, which may be coarse or fine. A coarse element is the one that covers a broad idea or prospective such as "address information". The fine granularity element presents a very specific and determined idea or prospective, such as the element of "street-name". The harmony in granularity ensures that the data and the knowledge components are presented at the same level of detail.

Figure 2.6 illustrates examples of positive and negative cases with harmony in representation, coverage and granularity. In summary, the mapping procedure matches the input elements with the knowledge entries. To allow the mapping process to be executed, the data and the utilized knowledge source should be harmonized.

Data Input	Knowledge Source	Harmony
$Input_1$ {**-50**} $Input_2$ {**33**} $Input_3$ {**-10**} $Input_4$ {**84**} . $Input_n$ {**9999**}	**$Domain_1$ (Real-Integer Numbers)** …………-9999, -9998,-9997,-9996, ,………… 0, 1, 2, 3, 4, 5, 6, 7, 8, 99999999………………	*Harmonized*
	$Domain_2$ (Integers in Words) ……………"Negative nine hundred",……..… "Negative one", "Zero" ,"One", "Two",……… …………….. "Nine hundred fifty four",……..	*Non-harmonized in representations*
	$Domain_3$ (Natural Numbers) 1, 2, 3, 4, 5, 6, 7, 8, 9, 10,11,12,13,14,15,16, ………………….., 99999999 ………………….	*Non-harmonized in coverage*
	$Domain_4$ (Sets) ………,{-10 - -1}, {0-9}, {10-19}, {20-29}, ……. ……………….{99990-99999}………………….	*Non-harmonized in granularities*

Figure 2.6: The harmony between the data and the knowledge

2.5.2 Mining Procedure

The mining procedure is the core process of the semantics extraction. This procedure operates on the matched elements and extracts the final output. The design of mining procedure follows the syntax of the knowledge source, as the mining procedure operates over its tags and the relationships. Also, the form of the mining procedure has to adhere to the problem on hand and the desired output.

The most commonly implemented mining procedures in the literature are *semantic similarities* and *rule-based flooding*, which are both mainly utilized with the ontologies form of knowledge.

2.5.2 (a) Mining Procedure through Flooding

Flooding procedure is an algorithm for searching a tree to identify a set of concepts related to the input one(s). In the semantics extraction process, *the flooding procedure* is executed over the hierarchical structure of the ontology. Over the knowledge hierarchy, the flooding procedure transfers from a given concept (i.e., a vertex in the structure) to another, sequentially throughout the hierarchical relationships till reaching a dead-end.

The *rules* attached with the flooding procedure determine the transferring form and direction of the flooding process. Generally, flooding can be implemented in two directions: bottom-up and top-down. In the bottom-up approach, the procedure transfers from one vertex to another up to the root vertex, as illustrated in the previous example of Figure 2.5. In the top-down approach, flooding starts at the upper level and continues down to some leaf vertex. The algorithm for flooding in the top-down approach is given in Algorithm 2.1.

Algorithm 2.1: Top-down Rule-based Flooding

```
FLOODING (T,v)
Begin:
1.  If v is leaf
2.      Output {o}← v
3.  End If
4.  Else
5.      For all the edges e in the out-going edges(v)
6.          v’ ← vertex (v,e)
7.          FLOODING (T,v’)
8.      End For
9.  End else
End
```

In Algorithm 2.1, the inputs to the flooding procedure are: tree (T) which corresponds to the hierarchy structure of the ontology and an input vertex (v) which corresponds to a given concept. The process starts at line 1, by checking if the active vertex (i.e., the vertex under exploration) is a leaf. If true, then this vertex is added to the output set in line 2. If the vertex is not a leaf, its connected edges are retrieved in line 5. In line 6, for each of the connected edges, the node that is on the other side of that edge is extracted and assigned as the active vertex. In Line 7, the flooding procedure is carried on for each new activated vertex. Subsequently, the overall process in Algorithm 2.1 gathers the leaves that can be reached from the initial input vertex (v) in the output set.

Example

An example of the discussed flooding procedure is illustrated in Figure 2.7. Given that the initial active vertex is $concept_2$, then the output is the leaf vertices $concept_8$ and $concept_9$.

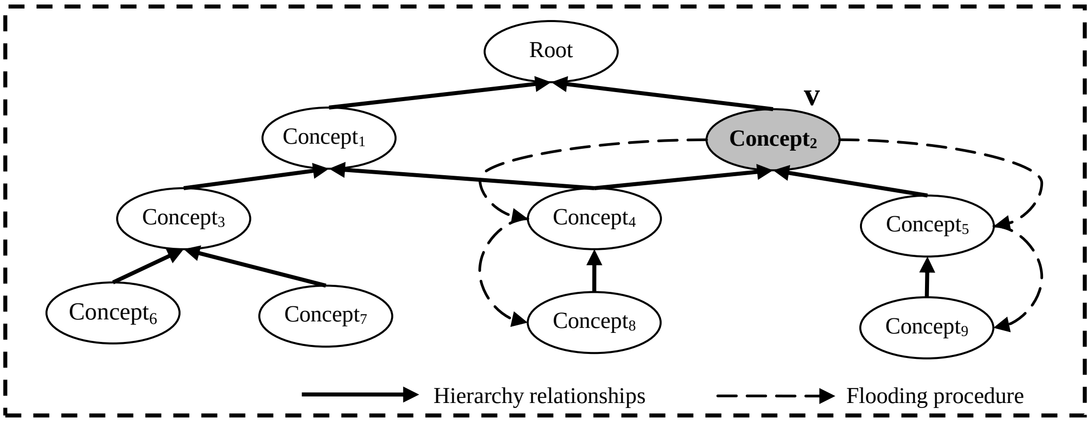


Figure 2.7: Example of a top-down flooding procedure

2.5.2 (b) Mining Procedure through Semantic Similarity

The *semantic similarity methods* measure the similarity and relatedness between a pair of concepts over a given ontology. Generally, there have been several methods to compute semantic similarity. Those methods can be categorized into edge-based, information content-based and feature-based. The edge-based and feature-based methods can be used with the semantics extraction process as they depend on having a knowledge source only. However, the information content-based methods require a corpus of textual data.

The *edge- based method* measures the relatedness between the input concepts based on the number of the intermediate edges/relationships between the concepts to be measured. Generally, the more edges there are and the greater the distance between the measured concepts, the lower the similarity. The *feature-based method* measures the similarity between the input concepts based on certain features, such as their definition or *glosses*. For example, Lesk (1986) measures the similarity between two concepts based on the number of common words in their glosses/definitions. The more common words there are, the more similar the input concepts are.

Comparative Study

Based on the comparison conducted by Petrakis et al. (2006), *Leacock and Chodorow's* (1998) method, gives the highest performance among the methods that can be executed based on knowledge source only. The comparative study, which is summarized in Table 2.1, is conducted over a set of concept pairs that is independent from any application based on WordNet and Mesh ontologies. The correlation, which is the basic factor for the comparison study, is a measure of how well the results obtained compare with the ground truth given by humans. A similar experimental study conducted by Budanitsky and Hirst (2001) reaches similar conclusions. The *Leacock and Chodorow* (1998) method is described through an example below.

Table 2.1: Evaluation of semantic similarity measures as provided by Petrakis et al. (2006)

Method	***Type***	***Correlation WordNet***	***Correlation Mesh***
Rada. (1989)	Edge	0.59	0.50
Wu and Palmer (1994)	Edge	0.74	0.67
Li et al. (2003)	Edge	0.82	0.70
Leacock and Chodorow (1998)	**Edge**	**0.82**	**0.74**
Richardson et al. (1994)	Edge	0.63	0.64
Tversky (1977)	Feature	0.73	0.67
Petrakis et al. (2006)	Feature	0.74	0.71
Rodriguez et al. (2003)	Hybrid	0.71	0.71

Example

Given the input concepts of "Grass" and "Acrogen" that have been identified using a mapping procedure, the Leacock and Chodorow Equation as given in Equation 2.1 and a part of WordNet is given in Figure 2.8.

$$sim(c_1, c_2) = -\log \frac{shortestLen(c_1, c_2)}{2.TaxonomyDepth} \tag{2.1}$$

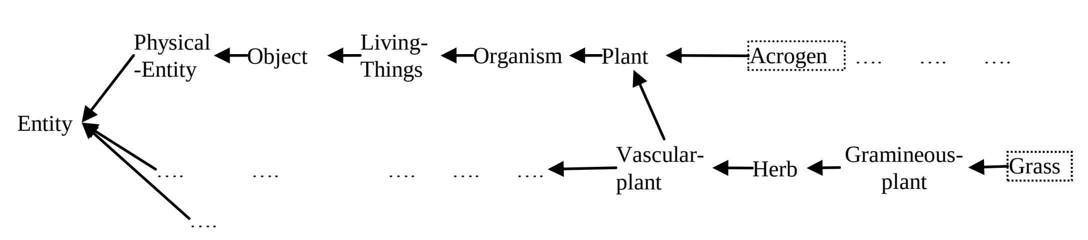


Figure 2.8: Part of the WordNet hierarchical structure

The similarity between the input concepts based on Leacock and Chodorow is measured as follows:

- Taxonomy Depth: the number of nodes from the root to the furthest vertex in the hierarchy. In Figure 2.8, the Taxonomy Depth, the distance between root node and “Grass”, is equal to **nine.**
- Shortest Length: the number of intermediate nodes in the shortest path connecting the input concepts c_1 and c_2. The shortest length between the concepts “Grass” and “Acrogen” in Figure 2.8 is **five.**
- Finally, the similarity can be calculated as -log(5/18). The final output is 0.55, which refers to the similarity between the concepts “Grass” and “Acrogen”.

As mentioned earlier, based on the Leacock and Chodorow measure, the more intermediate nodes there are and the greater the distance between the measured concepts, the lower the similarity. More examples of the similarities over the same knowledge source are given as follows:

- Similarity between “Plant” and “Acrogen” is 1.25.
- Similarity between “Plant” and “Grass” is 0.65.