People's Democratic republic Algeria

Ministry of higher education and scientific research
University of Abbes Laghrour khenchela

Faculty of sciences and technology

Department of Mathematics and Information Technology

Specialization: Web Security and Technology

Modeling and experimentation of a new measure of semantic similarity in a Conceptual hierarchy

Submitted by:

- ***TobbiSouhila***
- ***CherietAmel***

Supervised by:

- ***Dr. Djamel NESSAH***

BOARD OF EXAMINERS

- President: ....
- Supervisor: ....
- Examiner: ....

Dedication

In the Name of God, the Most Gracious, the Most Merciful.

All the Praise is Due to God Alone, the Sustainer of all the World.

I dedicate this dissertation:

- *First and foremost, to Dr. Nessah Djamel for all his efforts, and his support for us in every step of our work, we say all our thanks and appreciation to you .*
- *To my dear parents "Belkhir" and "Ghaoui Aicha" for their patience and assistance throughout my studies. You are our strength and our support in life, may God protect you.*
- *To my brothers, everyone in his name.... You are the true happiness and an example of sacrifice. Thank God you are my brothers.*
- *For every friend who was honest with us and for every wonderful friend who brought joy to our hearts, thank you.*

Thank you all.

Amel

Dedication

In the Name of God, the Most Gracious, the Most Merciful.

All the Praise is Due to God Alone, the Sustainer of all the World.

We dedicate this dissertation:

- *First and foremost, to Dr. Nessah Djamel for all his efforts, and his support for us in every step of our work, we say all our thanks and appreciation to you .*
- *To my dear parents "Moukhtar" and "Djeffali Betita" for their patience and assistance throughout my studies. You are my strength and support in life, may God protect you.*
- *To brothers and sisters and to my cousin "Imane" .. You are the true happiness and the example of sacrifice. Thank God you are in my life.*
- *For every friend who was honest with us and for every wonderful friend who brought joy to our hearts, thank you.*
- *And finally to that special person who had helped me and was a supporter of me, thank you.*

Thank you all. *souhila*

Résume :

Le concept de similarité entre les ressources d'information est aujourd'hui au cœur des études de recherches intenses.

L'objectif est de pouvoir mesures avec la précision la plus adéquat le degré de ressemblance entre ces ressources

Dans ce sens plusieurs mesure de similarité ont été développés, mais leurs efficacité reste sujet de critique et d'opinions pour apporter une amélioration nous présentant dans ce mémoire une nouvelle mesure de similarité sémantique dans les expérimentation montrant une nouvelle amélioration de la précision ressemblance enter les ressource documentaire

ملخص:

يعتبر مفهوم التشابه بين مصادر المعلومات في الوقت الحاضر موضوع الدراسات البحثية المكثفة.الهدف هو أن تكون قادرا على قياس درجة التشابه بين هذه الموارد بأكبر قدر من الدقة بهذا المعنى ، تم تطوير العديد من تدابير التشابه.

لكن كفاءتها تظل عرضة للنقد والآراء لإحداث تحسين. في هذه الأطروحة، نقدم مقياسًا جديداً للتشابه الدلالي في التجارب التي تظهر تحسنًا جديداً في دقة التشابه بين الموارد الوثائقية.

Abstract:

The concept of similarity between information resources is nowadays at the heart of intense research studies.

The objective is to be able to measure with the most adequate precision the degree of similarity between these resources

In this sense, several similarity measures have been developed, but their efficiency remains subject to criticism and opinions to bring an improvement. In this thesis, we present a new measure of semantic similarity in experiments showing a new improvement of the precision of similarity between documentary resources.

TABLE OF CONTENTS:

General Introduction: 8

Chapter 01:Knowledge Managementin The Content of Intelligence Artificial Applications 10

Introduction: 11

2. Knowledge management : 11

2.1Knowledge managementdefinition : 11

2.2 Evolution of Knowledge Management: 11

2.3knowledge management important:... 12

3. Knowledge Management Process: 12

3.1 Discovery: 13

3.2 Capture: 13

3.3 Organization: 13

3.4 Assessment:.. 13

3.5 Creation: 14

3.6 Sharing: 14

3.7 Reuse/application: 14

4. The knowledge management cycle (KMC) model :.. 14

4.1 Knowledge Creating: 15

4.2 Knowledge Sharing: 16

4.3 KnowledgeStructuring: 17

4.4 Knowledge Using: 18

4.5 Knowledge Auditing: 18

5. Software Development and KM: 19

5.1 Knowledge in Software Organizations: 19

5.2 KnowledgeManagement in Software Engineering: 19

6. Data, information and knowledge : 20

6.1 data definition: 21

6.2 information definition: 21

6.3 knowledgedefinition : 21
7. Aspects of reasoning: 24
7.1 formal reasoning : 24
7.2 Procedural reasoning: 24
7.3 Analogical reasoning: 24
7.4 Abstraction and generalization: 25
8. KNOWLEDGE MANAGEMENT TOOLS: 25
8.1 Requirement of Knowledge Management Tools: 25
8.2 Tools available for Knowledge Management: 26
9.Knowledgemanagement systems (KMSs) 27
9.1 Enterprise-wide knowledge management systems: 28
9.2 Knowledge work systems: 28
9.3 intelligent techniques: 28
10.knowledgemanagement in the context of artificial intelligence applications: 29
10.1 Applying AI techniques to advance knowledge management: 29
Conclusion: 30

Chapter 2: Knowledge Representation Models and reasoning 31

Introduction: 32
2. Knowledge representation: 32
2.1 Definition of knowledge representation: 32
3. Models: 32
3.1semantic networks: ……... 32
3.2Logical Representation: 36
3.3 Conceptual graphs: 39
3.4 Description logics: 42
3.5 Ontology: 44
Conclusion: 46

Chapter 03: Similarity Measurement Approaches and Reasoning 47

Introduction: 48

2.Useof similaritymeasures: 48

2.1 similarity in the data analysis: 48

2.2similarity in pattern recognition: 49

2.3similarity in case based reasoning: 49

3.Semantic similarity: 49

3.1 Notion of Semantic Similarity: 49

3.2 Semantic similarity measurement: 51

4.1 Ontology based measures: .... 51: .

4.2Information ContentBased method:.. 53

4.3 Hybrid methods : 54

4.4 feature based measures: 55

4.5 statistical methods: 55

5.Comparison and Evaluation 55

Conclusion: 57

Chapter04: Proposal of a new measure 58

Introduction: 59

2 Description of the approach: 60

2.1Formalization of our approach 63

2.2Intrinsic properties of the Hs measure 64

3- Implémentation: 67

3.1 Experiment in the WordNet hierarchy: 67

3.2Used languages and libraries:... 69

3.3Hierarchy generation algorithm: 69

3.4Our project workflow: 72

3.5experiments in a generic hierarchy : 74

3.6Discussion of results: 77

Conclusion:.. 77

General Conclusion: 78

Bibliography: 80

TABLE OF FIGURES:

Figure 1.1 Knowledge Management Life Cycle Model (Sağsan, 2006) 15
Figure 1.2: Data, information and knowledge pyramid (Nürnberger) 20
Figure 1.3.types of knowledge (Gupta, 2022) 21
Figure1.4 types of knowledge management systems (2009) 27
Figure2.1: example of semantic network 34
Figure 2.2: Architecture of knowledge bases in LDs (Franz Baader, 2003) 43
Figure 2.3: Knowledge base and interpretations in DLs (nessah, 2013) 43
Figure 3.1; Different notions of similarity (Schwering, 2008) 50
Figure3.2: categories of Semantic similarity measures (S. Elavarasi, 2014) 51
Figure3.3: Example of an ontology extract (T. Slimani, 2006) 52
Figure4.1: Conceptual hierarchy example 61
Figure 4.2: Sub-hierarchy for the first meaning of the word "red" in WordNet 68
Figure4.3 : Hierarchy generator function 71
Figure4.4: Children array completion function 70
Figure4.5: Append hierarchy concept root hypernyms (parents) 72
Figure4.6: Web application home page.. 73
Figure4.7: Web application result page.. 72
Figure4.8: Information Content Implementation 72
Figure4.9:Our proposed measurement implementation 74
Figure4.10: Trend of Hs and Wu-P measurements 76
Figure4.11: Trend of Hs and D.Lin measurements 7

LIST OF TABLES:

Table 3.1: Comparison of Different Semantic Similarity Measures**Erreur ! Signet no**
Table 4.1: Hs similarity applied to the hierarchy in Figure4.1 68
Table 4.2: Data and measures of Hs, Wu-P and D.Lin similarities 75
Table 4.3 : Accuracy number of identical values...76

General Introduction:

The quantity of digital documents available on the web is constantly growing, the traditional methods of searching for relevant information in a given context have become complex, inefficient and only rarely meet the needs expressed.

This state is mainly due to the neglect of the documents content, only the syntactic and morphological aspects are considered, and this has led to searches based on indexing terms without dealing with their content. **(lyd et all,2017)** As a result, the user wastes a lot of time searching and consulting documents that do not meet his requirements.

It then became necessary to research for powerful computer tools that would make it possible to select in a relevant way the documents which constitute an adequate response to its need, to classify the results, to preserve them and to present them in an intelligent way according to their degree of relevance.**(xav,2006).**

Our work, which is part of the research field in the semantic web, is a contribution that focuses on inter-document similarity measures; indeed, several applications and services in the current web are based on the semantic web to better serve their real and potential users. The use of semantics to take into account the contents of documents and not only the syntactic form has become a common problem, which has led to many research works including the appearance of similarity measures.

It is obvious that comparing documents based on the syntax of the concepts and terms contained in two documents for example would not give the desired answers. For example, if we are interested in the collection of computer products and we find in a document the word 'mouse' and another document the same word while the latter is interested in mammalian animals, it is clear that one or the other does not meet our expectations.

The use of ontology's to describe semantics is one of the solutions, it would allow to model a set of knowledge in a given domain, in particular we are interested in hierarchies like the WordNet lexicon to apply similarity measures.**(kao et all,2013)**

As for the existing approaches, described in the previous chapter, they can be seen as heuristics.

The heuristic is commonly used in the field of artificial intelligence, unlike algorithms, it is drawn from experience or analogies, rather than a scientific analysis too complex, the heuristic may consist in giving the idea of a proof, it is a reasoning that appeals to intuition or is based on the study of favorable cases

The one we propose combines approaches that have already been verified and used, the objective is to bring an improvement by a combination of two approaches belonging to two main categories:

- WU-Palmer's approach which is part of the path length based approaches.
- D.lin's approach which is part of the approaches based on the information content.

We will see in the experimental results that our approach brings more precision in these measures by producing more distinct values, and thus establishes a certain bijection between the set of concepts and the set of computed values.

We point out here that the morphological treatments of the terms that represent the concepts in the documents is not part of this work, the concepts are assumed to be extracted and preprocessed.

Chapter 01:

Knowledge Management in the Context of Intelligence Artificial Applications

Introduction:

Since the appearance of primitive human before thousands years ago, he has acquired a set of knowledge and in various ways, including communicating with others and communicating with the elements and sounds of nature, and since then he has stored information.

With the development of mankind and technology, there were huge amounts of information and knowledge, and it became difficult to control and manage them, so man tried to search for a way to facilitate this process, which would be useful in collecting information, discovering, creating, storing, sharing and restoring it when needed, so the name of this process is knowledge management.

Many specialists in the science of knowledge management have contributed to the emergence of the concept of knowledge management, including Peter Ducker and Donald Marchland, "Darker emphasized the increasing importance of information and explicit knowledge as organizational resources in institutions.

The beginning of the term knowledge management dates back to the beginning of the eighties of the last century, as it was the final stage of the hypotheses related to the development of information systems. In this chapter, we will talk more about what is knowledge management?

2. Knowledge management:

2.1Knowledge management definition:

Knowledge management (KM) or knowledge sharing in organizations is based on an understanding of knowledge creation and knowledge transfer also the (KM) is a discipline that crosses many areas such as economics, informatics, psychology and technology. KM is seen as a strategy that creates, acquires, transfers, consolidates, shares and enhances the use of knowledge in order to improve organizational performance and survival in a business environment .(Raz, 2020)

2.2 Evolution of Knowledge Management:

The three practices that have brought the most contents and energy to KM are information management, the quality movement and the human factors/human capital movement. Information management developed during the seventies and eighties and is usually

understood as a subset of the larger information technology and information science world. IT is a body of thought and cases that focus on how information itself is managed independent of the technologies that house and manipulate it. It deals with information issues in terms of valuation, operational techniques, governance, and incentive schemes.
In broad terms, KM shares information management's user perspective, a focus on value as a function of user satisfaction rather than the efficiency of the technology that houses and delivers the information.(Gha, 2009)

2.3knowledge managementimportant:

Knowledge is the heartbeat of every decision made within your organization. Product development, service improvement, customer engagement, and the entire company operation couldn't function without the knowledge your team hold(Chr, 2021).
As a logical next step, you need to ensure that the knowledge available in your organization is distributed throughout the organization. This is where the knowledge management process comes into play.
To organize this process in an accessible and efficient way, you need a knowledge management system, which in turn allows you to:

- Keep the knowledge documented, structured, and easy to find
- Standardize processes across the organization
- Make training materials accessible and well organized
- Save your employees' time they would spend searching information and managers' time explaining the same things over and over again
- Mitigate 'bus factor' risks when employees dealing with particular knowledge areas are having vacation or sick leave (Chr, 2021)

3. Knowledge Management Process:

A knowledge management process is the way in which a business manages knowledge, including its capture, storage, organization, verification, security, distribution, and use. This process, when effective, informs accurate and beneficial decision-making, stimulates collaboration and innovation throughout the organization, and enhances internal and external communication and efficiency A knowledge management project should support the activities that comprise a knowledge process[1] ,According to Staab a knowledge process involves the following steps:

3.1 Discovery:

The process of knowledge management begins with discovery. Knowledge discovery is the process of extracting information from data that can be useful to your organization's strategy, operations, communication, and relationship development. Using data mining to identify patterns, trends, or correlations within large sets of transactional or customer relationship data is an example of discovery. [1]

3.2 Capture:

Knowledge capture is acquiring the knowledge your organization already possesses—within individual employees, teams, documents, or processes—as well as external knowledge, so it can be documented, communicated, and shared to the benefit of your business. Conducting an audit of your existing documentation and encouraging content creation in knowledge areas where you have gaps are key to capturing knowledge.[1]

3.3 Organization:

Knowledge organization means describing, classifying, categorizing, and indexing information so it can be easily retrieved, navigated, reused, and shared among employees, teams, and other critical users. The right knowledge management system can help you sort and segment knowledge so information is readily accessible by the people who need it most.[1]

3.4 Assessment:

For knowledge that drives useful business decisions, inspires collaboration and innovation, and improves internal and external processes, you need to ensure that knowledge is validated and validated. That means integrating processes that ensure any information your organization intends to apply is accurate, complete, consistent, and up to date. Automatic validation features within your knowledge management system, as well as regular reviews by your internal experts, are critical. [1]

3.5 Creation:

Knowledge creation is when individuals or teams within your organization add what they have learned—through practice, process navigation, internal and external interactions, independent research, and other experiences—to the organization's collective knowledge. That individual knowledge can then be shared, reused, and applied, as well as expanded upon by future knowledge seekers[1]. Establishing content creation guidelines and regularly communicating the value of new, relevant content - as well as praising/rewarding completion - can help you build a culture of knowledge creation within your team

3.6Sharing:

Knowledge sharing encompasses both making knowledge available to those who actively seek it within your organization and directly communicating appropriate knowledge to a user who could potentially apply it for the benefit of your business. Team leaders should regularly encourage, and possibly incentivize, knowledge sharing. [1]

3.7 Reuse/application:

Knowledge reuse or application is when an individual or team can take captured (and organized/assessed) knowledge and apply that knowledge to enhance efficiency, improve business operations, complete a strategic task, communicate more effectively with colleagues or customers, etc. A documented lesson learned from one employee's complex customer interaction may streamline the process for a colleague in a similar situation, for example.[1]

4. The knowledge management cycle (KMC) model :

In accordance with knowledge management literature, five basic processes can be considered by managing knowledge. These can be defined as creating, sharing, structuring, using, and auditing in turn that is called "knowledge management life cycle" model. This model makes us to understand knowledge management processes in hierarchical order. Each model is explained in the following paragraphs.(Sağ, 2006)

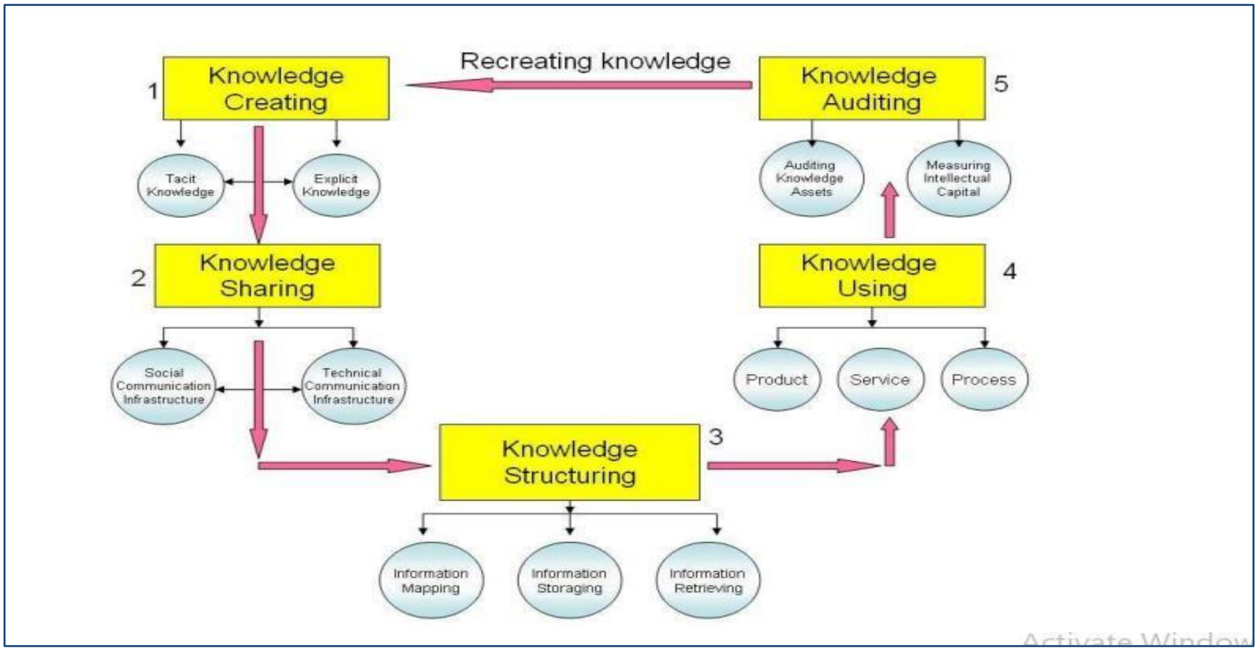


Figure 1.1 Knowledge Management Life Cycle Model **(Sağ, 2006)**

4.1 Knowledge Creating:

The first stage of managing organizational knowledge requires entering the 'knowledge kitchen'.

In other words, exploring knowledge creating stage where can be processed in organization leads us to focus which individual, group, and department on. Because if knowledge can not be created in organization; neither sharing nor auditing knowledge can be carried out.

There are too many knowledge creators in knowledge kitchen due to the fact that organization can not create collective knowledge by itself. Thus, organizational participants create knowledge through their intuition, ability, skills, behaviors, and work Experiments.

'Key players, departments and their interactivity can play a critical role in creating knowledge in organization' .

It is gained through formal education or structured study and codifies, stores, hierarchy of databases and accesses with high quality, reliable, fast information retrieval systems. Therefore, explicit knowledge is easy to structure and retrieval.

Another form of knowledge is tacitness which completely individual and collective. Tacit knowledge is a personal form of a knowledge, which individuals can only obtain from direct experience in a given domain. It is held in a non-verbal form, and therefore the holder can not provide a useful verbal explanation to other individual. Nonaka emphasizes two dimensions of tacit knowledge. These are technical and cognitive. Technical dimensions covers the kind of informal personal skills of crafts often referred to as «knowhow». 'Knowing-how' is characteristic of the expert, who acts, makes judgments, and so forth without explicitly relecting on the principles or rules involved. Cognitive dimension consists of beliefs, ideals, values, schemata, and mental models.

In the knowledge kitchen, tacit knowledge is transferred by using for organization's products/services and work processes and this conversion gives rise to competitive advantage between participants in business units.(Sağ, 2006)

4.2 Knowledge Sharing:

The second important stage of knowledge management life cycle is knowledge sharing. ,emphasizes the ways and tools for effective knowledge sharing as follows:

- formal social communication network.
- informal social communication network.
- teamwork.
- communities of practices.
- organizational learning.
- rumors and.
- formal structured technological communication networks (e-mail, mobile communications ,teleconferences, videoconferences, etc.).

there are two types of knowledge sharing :

- Constructing Social Communication Infrastructure.
- Constructing Technical Communication Infrastructure.
  (Sağ, 2006)

4.2.1 Constructing Social Communication Infrastructure: Written communication includes memos, letters, electronic mail, fax transmissions, organizational periodicals, notices placed on bulletin boards, or any other device that is transmitted via written words or symbols. Nonverbal communication entails body movements, the intonations or emphasis we give the words, facial expressions, and the physical distance between the sender and receiver . Knowledge management prefers all three forms of organizational communication because the effective knowledge management system requires all forms of knowledge such as written/verbal, explicit/tacit, audio/visual in organization. Successful knowledge management strategies entail particularly grapevine communication networks on the grounds that these channels are more persuasive and reliable rather than formal communication channel because of supporting managers. (Sağ, 2006)

4.2.2 Constructing Technical Communication Infrastructure:

Technical communication infrastructure refers to information and communication technology. Participants can share their expertise knowledge through e-mail, in-group computerized communication networks, databases, telephone conversations. Technical communication infrastructure which is known as formal communication networks provide in sharing, structuring, classifying and organizing explicit/tacit knowledge in the environment.(Sağ, 2006)

4.3 Knowledge Structuring:

After establishing a sound knowledge sharing infrastructure system; the structure of data, information and knowledge should be stored in the organization's database for future needs. The basis of structured knowledge is the categorization, organization, coding, analysis and reporting of information so that information about what the organization needs in the future can be retrieved. This means that mapping, storing and retrieving information are three important components of knowledge structure. First, information related to identifying sources of organizational information and information known to participants is mapped. The second is information storage, including the storage of knowledge such as databases, data warehouses, and information centers, and an electronic environment that displays organizational memory. Third, the most important factor in constructing knowledge is called information retrieval

4.4 Knowledge Using:

Organizations use knowledge for three reasons:

1) Knowledge can be used for determining organization's work processes and making strategies for sustainable competitive advantage.
2) Knowledge can be used for designing and marketing product.
3) Knowledge plays a critical role of organization's services quality (Sağ, 2006)

   Organizational routines refer to the development of task performance and coordination patterns, interaction protocols and process specifications that allow individuals to apply and integrate their specialized knowledge without the need to articulate and communicate what they know to others(Sağ, 2006).

4.5 Knowledge Auditing:

Knowledge auditing means what amount of knowledge can be used in organization's products, services and processes. This knowledge management life cycle stage refers to the capacity of information processing in organizations. In other words, what amount of information and knowledge are created, shared, stored, and used in organization in a certain time helps us to determine information capacity in organizations(Sağ, 2006). The knowledge audit provides value when company is doing one or more of the following:

- Devising a knowledge-based strategy.
- Architecting a knowledge management blueprint or roadmap.
- Planning a build a knowledge management system.
- Planning research and development.
- Seeking to leverage its 'people assets'.
- Facing competition from knowledge intensive competitors that are far ahead on the learning curve.
- striving to strengthen its own competitive weakness
- looking for direction for planning a market entry or exist strategy (Tiw, 2000)

When we look at the perspectives of knowledge management application in organization, this life cycle model is to encourage Chief Knowledge Officer how knowledge management should be succeeded. Obviously, the answer is knowledge management. Because it is not only a new style of management, but also it is a new business model that focuses on knowledge-intensive works in organizations(Sağ, 2006).

5. Software Development and KM:

In software development, knowledge management is seen as a process that spans the different stages/phases of the SDLC and involves sharing tacit knowledge and transforming it into explicit knowledge, which is then transformed through personal experience and other factors such as beliefs, opinions, and values into Tacit knowledge. Therefore, the individual is the most important player in knowledge management, creating, sharing, enhancing, enhancing and justifying knowledge of the organizational framework through social and collaborative work tasks among people in the organization. KM is expected to be an integral part of any software development environment.

5.1 Knowledge in Software Organizations:

A software organization has many different needs related to knowledge. These needs can be viewed from a business and from a skills and practice perspective. From a business perspective, the main needs are to produce better, faster, and cheaper software and to make better decisions. (Mik et all., 2003)

Software organizations have and require vast amounts of knowledge to support the business objectives for which technology, process, project, product, and domain knowledge are the most critical areas.(Mik et all., 2003)

5.2 Knowledge Management in Software Engineering:

“Software engineering is a knowledge intensive business and as such it could benefit from the ideas of knowledge management. The important question is, however, where does knowledge reside in software engineering? It is clear that software engineering involves a multitude of knowledge-intensive tasks: analyzing user requirements for new software systems, identifying and applying best software development practices, collecting experience about project planning and risk management, and many others”(Bir et all., 1999)

6. Data, information and knowledge:

Knowledge, information and data are key words and also key concepts in knowledge management, intellectual capital and organizational learning. Here we will define these concepts.

Through Raphael Kaburu :'data ,information and knowledge .putting together the three concepts ("data", "information" and "knowledge") as applied herein, gives the impression of a logical hierarchy : information is a set of data and knowledge comes from information collection .this is a fictional story."(Edo, 2014)

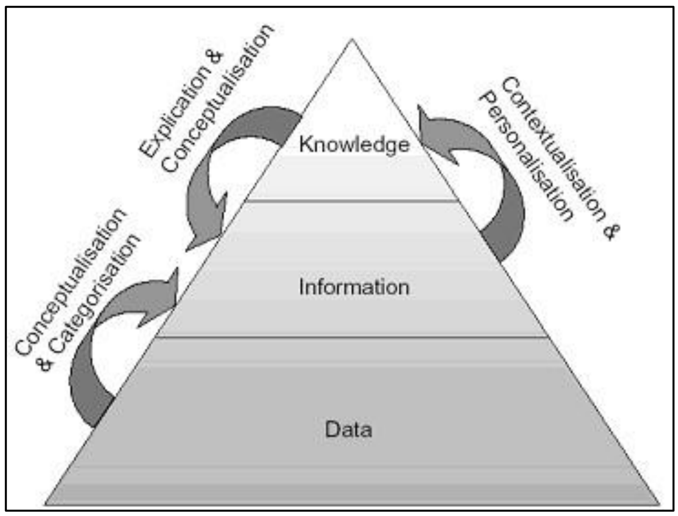


Figure 1.2: Data, information and knowledge pyramid **(Nür et all., 2011)**

Let's see what does means every one:

6.1data definition:

Data can be defined as mere, unformatted or raw facts. This could in inform of measurements, statistics, numbers or alphabets (Ala et all. , 2001). Data in itself is usually meaningless as it has no meaning without being interpreted (Bou et all., 2002). Examples of data include "101010", "CUO", "13031305". (Edo, 2014)

6.2 Information definition:

Information can be simply defined as processed data. It is inference gained from data (Bouet all, 2002). Information can also be defined as a set of related data, which can be further interpreted and put into relevant context (Ala et all., 2001).(Edo, 2014)

6.3 Knowledge definition :

According to a dictionary definition, knowledge is "the fact or condition of knowing something with familiarity gained through experience or association; acquaintance with or understanding of a science, art, or technique; the fact or condition of being aware of something" (Mer, 2001).(Edo, 2014)

According to (Non et all., 1996) knowledge can be defined as a belief which is held to be justifiable and true. Some authors define knowledge as the application of information (Bou et all., 2002).(Edo, 2014)

6.3.1 Classification of knowledge:

There are 7 types of knowledge:

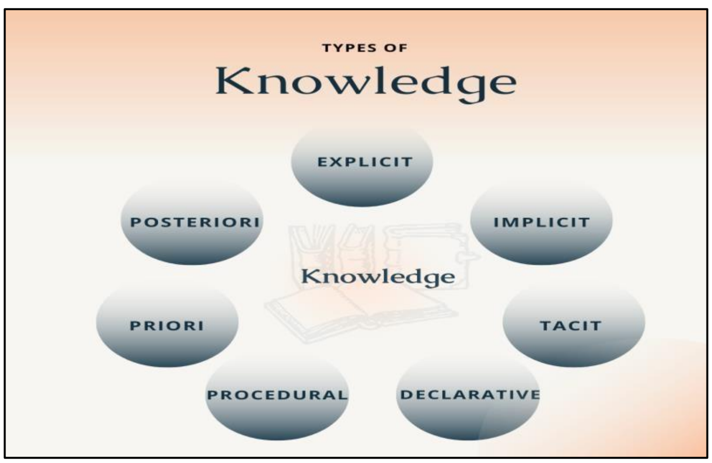


Figure 1.3.types of knowledge **(Gup, 2022)**

6.3.1.1 Explicit knowledge:

Explicit knowledge is knowledge that is straightforwardly expressed and shared between people. Explicit knowledge exists in the form of data that can be processed and stored, organized and interpreted. This type of knowledge is highly shareable between employees because it can be captured in a knowledge base or as part of your Knowledge Management strategy.(Pra, 2022)

Example: Information from tangible sources, such as books, newspapers, or articles.

6.3.1.2 Implicit (or tacit)knowledge :

Implicit knowledge or 'tacit knowledge' is the type of knowledge that is complex to transmit to another individual using verbalizing or writing it down. It can be well-defined as experiences, culture, ideas, and skills that the public has but are not classified and may not essentially be simply articulated.(Pat, 2020)

Examples:

Implicit: Information that does not originate in a tangible form, but can be transferred into tangible form, such as a dictation of a speech or an experience.(Raz, 2020)

Tacit: Information that's difficult to capture tangibly, such as your perception of an experience or a feeling after a big event.(Raz, 2020)

6.3.1.3 Declarative knowledge:

One such vertical that is defined is declarative knowledge. It refers to static facts or information stored in a database. This information can be conceptual, propositional, or even descriptive knowledge that describes subjects, things, events, outcomes, and their attributes about each other. (Pat, 2020)

Example**:** Declarative knowledge can be thought of as information about various categories of facts, such as world history, mathematical operations, and so forth. (Mai, 2022)

6.3.1.4 Procedural knowledge:

Procedural knowledge can be simply stated as knowing how to do something. It is defined as the knowledge attained by practicing or exercising a task or a skill. It is also called practical knowledge, imperative knowledge, or task knowledge.(Pat, 2020)

Now that we understand what procedural knowledge is, let's look at some areas where it can be highly valuable.

- **Intellectual Property:** Procedural knowledge, in intellectual property law, is tightly held information about a company's technology that could be used to create products and services of commercial value.
  Legally speaking, procedural knowledge is part of an organization's intellectual property and the company has the right to license it through patents and trademarks.(Pat, 2020)
- **Artificial Intelligence:** In artificial intelligence, procedural knowledge is knowledge retained by an intelligent entity. In this case, a mobile robot that's AI system-based is given specific instructions on navigation or actionable steps about pathways.

The inbuilt algorithm will then have to discover how to use this given resource to complete complex tasks at hand.(Pat, 2020)

Procedural Knowledge in Education:

Examples of procedural knowledge in the classroom include the impact of procedural knowledge that could help students to learn concepts better and deeper. In a way, procedural knowledge is part of prior knowledge. As a procedural knowledge and educational psychology example, in mathematics education when students first learn to count, they make use of their hands and fingers.(Pat, 2020)

- **Psychological Impact:**

As per many cognitive psychologists, individuals who learn procedural knowledge are sometimes unaware they are learning.(Pat, 2020)

6.3.1.5 Priori knowledge:

If a person can know that some statement is true just by knowing the meaning of the words in the statement, or just by contemplating what the statement means, then it is an example of a priori knowledge. The term a priori is italicized because it is originally a Latin phrase. In Latin, a priori means "from the former" or "from the one before.(Pon, 2021)

6.1.3.6 Posteriori knowledge:

A posteriori knowledge is knowledge that a person has learned from their experiences. In this way, it is the opposite of a priori knowledge.(Pon, 2021)

7. Aspects of reasoning:

The reasoning is a sequence of statements conducted in function of a goal. The goal can be a demonstration, elucidation, interpretation, or explanation, etc. .

The nature of chaining is an important characteristic of reasoning which in general requires backtracking. The model of reasoning is inherent to the knowledge on which it operates, thus we canon which it operates(Nes, 2013), so we can distinguish:

7.1 formal reasoning:

- which materializes through the syntactic manipulation of structures symbolic, logical reasoning is one of them.(Nes, 2013)
- In formal reasoning we're talking about things like algorithms, systematic methods that allow you to produce a correct solution to a problem, assuming that a solution even exists. Those formulas are examples of algorithms. And a formal reasoning would also include rules of logic, and syllogisms, syllogisms being those logical arguments where you're given two premises with a conclusion and you have to see, «Are the premises and conclusion valid? Does the conclusion logically follow from the premises?»And of course, you know, with any logical reasoning puzzle, we may have certain biases that lead us astray and perhaps cause us to make illogical conclusions or logical fallacies. [2]

7.2 Procedural reasoning:

Knowledge, its modes of use, and reasoning are explained in algorithms or reasoning are explained in algorithms or automata. The mechanisms of mechanisms of attachment of demons in frame systems highlight such reasoning. (Nes, 2013)

7.3 Analogical reasoning:

is a kind of reasoning that is based on finding a common relational system between two situations, exemplars, or domains. When such a common system can be found, then what is known about one situation can be used to infer new information about the other. The basic intuition behind analogical reasoning is that when there are substantial similarities between situations, there are likely to be further similarities. This article describes the processes involved in analogical reasoning, reviews seminal research and recent developments in the field, and proposes new avenues of investigation. (Gen et all, 2012)

7.4 Abstraction and generalization:

Are the processes of facilitating a specific problem to help designers solve problems efficiently? Abstraction and generalization reduce complexity and increase creativity. Both abstraction and generalization guide designers to focus on the key factors of a problem towards producing a broader solution perspective.. In addition, the aspects of abstraction and generalization advantages, their implementation in the design process, safety constraints and comparisons between abstraction and generalization are also reviewed. A case study of an aircraft component is used as the example in conducting abstraction and generalization in the safety approach.(Kha et all., 2016)

8. KNOWLEDGE MANAGEMENT TOOLS:

Knowledge management tools are any tools used to effectively gather and distribute both internal and external knowledge for your business. While available individually, most knowledge management tools are included as features in knowledge management software.[3]

8.1 Requirement of Knowledge Management Tools:

The tools for KM are focused on assimilation, comprehension, and learning of the information by individuals who will then transform data and information into knowledge. Thus the visible part of knowledge, what the literature calls explicit as opposed to the tacit dimension of knowledge, is only information regardless of the amount of the other individual knowledge embedded into it(Gha, 2009) . These reasons are :

8.1.1 Facilitate information contextualization:

To facilitate information contextualization, metadata on its characteristics and integration within a specific environment must be attached to it before storing. This facilitates better retrieval and management for the knowledge seeker.(Gha, 2009)

8.1.2 Intelligently transfer information:

Information transfer must occur by taking into account the user, the content, and the time of transfer. A tool that can optimize these three aspects can truly provide information according to the needs of the users, respecting one of the key functional foundations of KM(Gha, 2009).

8.1.3 Facilitate social interactions and networking:

Direct communication and verbal knowledge transfer through social interactions among individuals is the most natural aspect of knowledge sharing. A KM tool supports this social aspect and facilitates searching.(Gha, 2009)

8.1.4 Present a customized human-computer interface:

The tools also support interface customization and ease of use. The human-computer interface, ease of use and usability will drive intention to use and reuse the tools.(Gha, 2009)

8.2 Tools available for Knowledge Management:

A number of tools are available to support the functionalities and processes of KM, which are listed below:

8.2.1 Tools to access knowledge:

These tools provide access to explicit knowledge that can be shared and transferred through the enterprise information systems. For example, Convera is a tool used for retrieval ware. It works on powerful indexing systems to classify expertise based on both content and collaboration dynamics and networks within the enterprise.(Gha, 2009)

8.2.2 Tools for semantic mapping:

These tools are meant to quickly support presentation of information, analysis, and decision making. Ontology tools are also part of this category as they enable users to organize information and knowledge by groups and schemata that represent the organizational knowledge base.(Gha, 2009)

8.2.3 Tools for knowledge extraction:

These tools support structured queries and replies. They help mining text by interpreting relationships among different elements and documents. For example, clear Forest Text Analysis.(Gha, 2009)

8.2.4 Tools for expertise localization:

For example, Active Net maintains a continuous, real-time view of organizational activities. Active Net continuously discovers each person's work activity and business relationships by processing communications from such sources as documents, discussion databases, e-mail, instant messaging and digital workspaces.(Gha, 2009)

8.2.5 Tools for collaboration work:

These tools enable teams to globally share dedicated spaces for managing the project lifecycle; editing and publishing materials; conducting live discussions and interactions; and maintaining a repository of materials associated with every step of the process. Using QuickPlace, coworkers, suppliers, partners, and customers can communicate online immediately within a structured workspace created for that purpose.(Gha, 2009)

9. Knowledge management systems (KMSs)

Knowledge Management System (KMS) is a system for applying and using knowledge management principles to typically enable employees and customers to create, share and find relevant information quickly [4].

There are three major types of knowledge management systems:

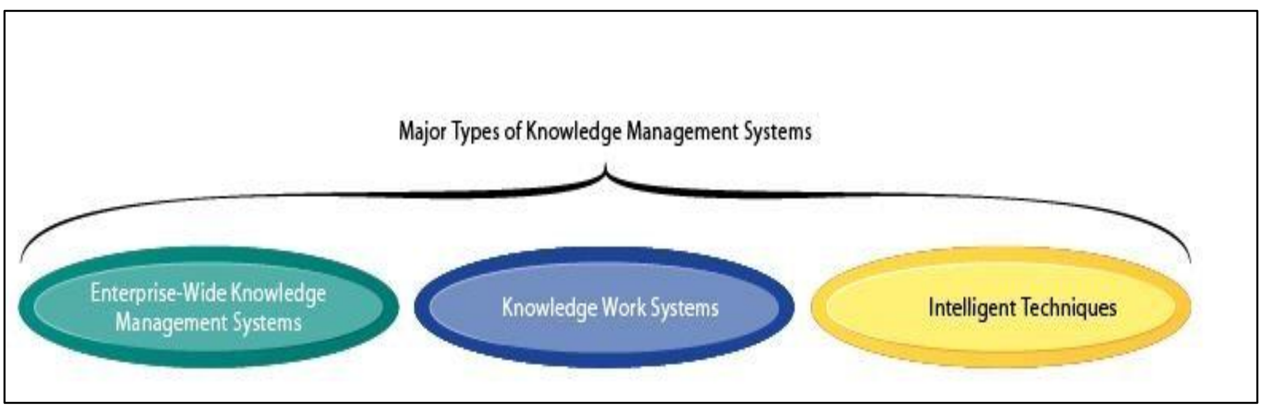


Figure1.4 types of knowledge management systems[5]

9.1 Enterprise-wide knowledge management systems:

Enterprise-wide knowledge management systems are firm wide efforts to collect, store, distribute, and apply digital content and knowledge. They use an array of technologies for storing structured and unstructured content, locating employee expertise, searching for information, disseminating knowledge, and using data from key corporate systems.[6]

9.2 Knowledge work systems:

Is a specialized system built to promote the creation of knowledge and to make sure that knowledge and technical skills are proper integrated into business. It helps the knowledge workers in creating and propagating new information and knowledge by providing them the graphics, analytical, communications, and document management tools.(Sha)

9.3 intelligent techniques:

Intelligence techniques may be used for:

- Capturing individual and collective knowledge and extending a knowledge base, using artificial intelligence and database technologies
- Capturing tacit knowledge, using expert systems, case-based reasoning, and fuzzy logic
- Knowledge discovery, or discovering underlying, hidden patterns in data sets, using neural networks and data mining
- Generating solutions to highly complex problems, using genetic algorithms
- Automating routine tasks, using intelligent agents [7]

10. Knowledge management in the context of artificial intelligence applications:

Knowledge management is an emerging area which is gaining interest by both industry and government. As we move toward building knowledge organizations, knowledge management will play a fundamental role towards the success of transforming individual knowledge into organizational knowledge. One of the key building blocks for developing and advancing this field of knowledge management is artificial intelligence, which many knowledge management practitioners and theorists are overlooking. (Lie, 2001)

10.1 Applying AI techniques to advance knowledge management:

In looking at ways for sharing knowledge, transforming individual knowledge into collective, organizational knowledge, and reincarnating organizations into «knowledge organizations», the field of artificial intelligence can help push these basic tenets of knowledge management.

One of the important areas of knowledge management is knowledge capture and representation. The knowledge engineering methodologies for building expert systems have applied knowledge acquisition techniques for eliciting the tacit knowledge from domain experts. In order to develop knowledge repositories in knowledge management systems for formally documenting knowledge in an on-line way, these knowledge acquisition techniques could be applied. Other AI techniques like intelligent agents can be used to help in the search and retrieval methods of knowledge in the knowledge management systems. Agents can be used to help in combining knowledge which would ultimately lead to the creation of new knowledge. The AIApplications Institute at the University of Edinburgh has developed an adaptive work system, using agent technology, to support knowledge management. Natural language and speech understanding front-ends as interfaces to knowledge management systems may be worthwhile AI techniques to apply in the coming years to the knowledge management field.

There are also techniques that could be used to assist in this process. Intelligent agents could be applied to analyze the knowledge, email, web pages, and the like and to disseminate appropriate summaries or individual pieces of information and knowledge to those who should best make use of it. Data mining and knowledge discovery techniques could also be employed to inductively look for trends, relationships, and possibly new knowledge and information from the organization's knowledge repositories. This is already being done effectively in the marketing and finance fields.(Lie, 2001)

Conclusion:

In conclusion, we conclude that knowledge management has a large and important role in managing systems and facilitating their work in various fields, as it is of great importance in technology today. Ithasalso a primary goal, which is to transform individual knowledge into collective knowledge, and to transform organizations into knowledge organizations. Artificial intelligence is one of the most important elements for developing this field of "knowledge management" and achieving its goals. For example, "intelligent agents" can be used to assist in research and information retrieval methods in knowledge systems.

Chapter 2:

KnowledgeRepresentation Models and reasoning

Introduction:

There is a big difference between man and machine, so that man is better at understanding, inferring and interpreting knowledge as opposed to machine.

But the development and speed the world is witnessing made us think of a faster way to make the machine can embody real world problems, understand them, and then solve them. And this method is Knowledge representation. So what is the knowledge representation?. At the beginning of this chapter, we will define the concept of knowledge representation, and then we will discuss its model and conclude with ontology.

2. Knowledge representation:

2.1 Definition of knowledge representation:

- Knowledge representation and reasoning is the area of Artificial Intelligence (AI) concerned with how knowledge can be represented symbolically and manipulated in an automated way by reasoning programs. More informally, it is the part of AI that is concerned with thinking, and how thinking contributes to intelligent behavior.(Ron et all., 2004)
- It is in charge of encoding information about the real world in such a way that a computer can comprehend it and use it to solve complicated real-world problems like diagnosing a medical condition or conversing with humans in natural language.[8]

-It's also a means of describing how artificial intelligence can represent knowledge. Knowledge representation entails not only storing data in a database, but also allowing an intelligent computer to learn from that knowledge and experience in order to act intelligently like a human.[8]

3. Models:

3.1 semantic networks:

3.1.1 Definition:

A semantic network is a graphic notation for representing knowledge in patterns of interconnected nodes. Semantic networks became popular in artificial intelligence and natural language processing only because it represents knowledge or supports reasoning.

These act as another alternative for predicate logic in a form of knowledge representation.(Mar, 2019)

- Semantic nets consist of nodes, links and link labels. In these networks diagram, nodes appear in form of circles or ellipses or even rectangles which represents objects such as physical objects, concepts or situations.
- Links appear as arrows to express the relationships between objects, and link labels specify relations.
- Relationships provide the basic needed structure for organizing the knowledge, so therefore objects and relations involved are also not needed to be concrete.
- Semantic nets are also referred to as associative nets as the nodes are associated with other nodes.

This representation consists of mainly two types of relations:

- IS-A relation (Inheritance)
- Kind-of-relation

Example:

Following are some statements which we need to represent in the form of nodes and arcs.

a- bob is a dog.

b- bob is owned by lisa.

c- Bob is black colored.

d- All dogs are animal.

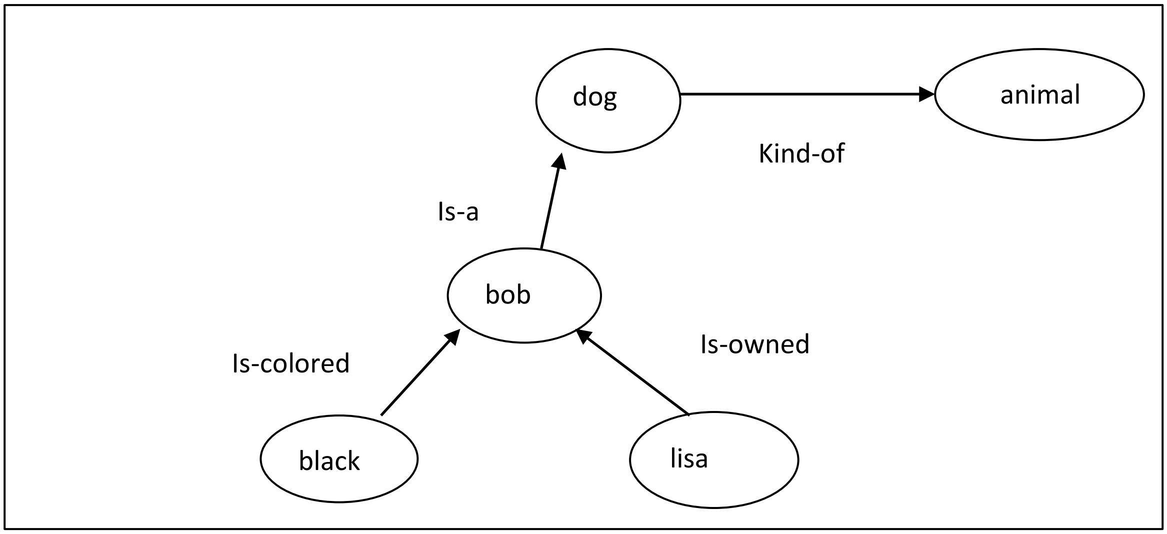


Figure2.1: example of semantic network

3.1.2 MainComponentsOf Semantic Networks:

- Lexical component: nodes denoting physical objects or links are relationships between objects; labels denote the specific objects and relationships
- Structural component: the links or nodes from a diagram which is directed.
- Semantic component: Here the definitions are related only to the links and label of nodes, whereas facts depend on the approval areas.
- Procedural part: constructors permit the creation of the new links and nodes. The removal of links and nodes are permitted by(Mar, 2019)destructors.

3.1.3 AdvantagesOf Using Semantic Nets:

- The semantic network is more natural than the logical representation;
- The semantic network permits using of effective inference algorithm (graphical algorithm)
- They are simple and can be easily implemented and understood.
- The semantic network can be used as a typical connection application among various fields of knowledge, for instance, among computer science and anthropology.
- The semantic network permits a simple approach to investigate the problem space.
- The semantic network gives an approach to make the branches of related components.

- The semantic network also reverberates with the methods of the people process data.
- The semantic network is characterized by greater cognitive adequacy compared to logic-based formalism.
- The semantic network has a greater expressiveness compared to logic. (Mar, 2019)

3.1.4 Disadvantages OfUsing Semantic Nets:

- There is no standard definition for link names
- Semantic Nets are not intelligent, dependent on the creator
- Links are not alike in function or form, confusion in links that asserts relationships and structural links
- Undistinguished nodes that represent classes and that represents individual objects
- Links on object represent only binary relations.
(Mar, 2019)

3.1.5 Six most common kinds of semantic networks:

3.1.5.1 Definitional networks:

These networks focus on and deal only with the subtype, or they are a link between a concept type and a newly formed subtype. A generalization hierarchy is a type of production network. It adheres to the inheritance rule when it comes to duplicate characteristic(Mar, 2019)

3.1.5.2 Assertion networks:

Are designed to assert propositions. Unlike definitional network, the information in an assertion network is assumed to be contingently true, unless it is explicitly marked with a modal operator.

Some assertion networks have been proposed as models of the conceptual structures underlying natural language semantics.(Mar, 2019)

3.1.5.3 Implicational networks:

Use implication as the primary relation for connecting nodes. They may be used to represent patterns of beliefs, causality, or inferences.

Implicational networks emphasize implication, they are capable of expressing all the Boolean connectives by allowing a conjunction of inputs to a propositional node and a disjunction of outputs.(Mar, 2019)

3.1.5.4 Executablenetworks:

Include some mechanism, such as marker passing or attached procedures, which can perform inferences, pass messages, or search for patterns and associations.
Executable semantic networks contain mechanisms that can cause some change to the network itself.(Mar, 2019)

3.1.5.5 Learning networks:

Build or extend their representations by acquiring knowledge from examples.The new knowledge may change the old network by adding and deleting nodes and arcs or by modifying numerical values, called weights, associated with the nodes and arcs.
The purpose of learning, both from a natural and AI standpoint, is to create modifications that enable the system to respond more effectively within its environment.(Mar, 2019)

3.1.5.6 Hybrid networks:

Combine two or more of the previous techniques, either in a single network or in separate, but closely interacting networks..
System is usually called hybrids if their component languages have different syntax... The most widely used hybrid of multiple network notations is the Unified Modeling Language (UML), which was by designed by three authors....who merged their competing notations.(Mar, 2019)

3.2 Logical Représentation:

3.2. 1. Définition:

A language with certain concrete principles that deals with propositions and has no ambiguity in representation is referred to as logical representation. Drawing a conclusion based on numerous criteria. Some important communication guidelines are laid out in this representation. It's made up of well-defined syntax and semantics that facilitate sound inference. Syntax and semantics can be used to translate each sentence into logics.[9]

3.2.2 Syntax:

Syntaxes are the rules which decide how we can construct legal sentences in the logic.
It determines which symbol we can use in knowledge representation.

How to write those symbols.[9]

3.2.3 Semantics:

Semantics are the rules by which we can interpret the sentence in the logic. Semantic also involves assigning a meaning to each sentence.[9]

Logical representation can be categorized into mainly two logics:

3.2.4 Propositional Logics:

The simplest and most abstract logic we can study is called propositional logic.

- Definition: A proposition is a statement that can be either true or false; it must be one or the other, and it cannot be both (Woo).
- Examples. The following are propositions:
  - the reactor is on;
  - the wing-flaps are up;
  - John Major is prime minister.

Whereas the following are not:

- Are you going out somewhere?
- 2+3

  [9]

3.2.5 Predicate logics:

A predicate is an expression of one or more variables determined on some specific domain. A predicate with variables can be made a proposition by either authorizing a value to the variable or by quantifying the variable.

The following is an example of predicates:

- Consider M(x, y) denote "x is married to y."[9]

3.2.6 Logical reasoning:

The reasoning on logical representations, will consist in a sequence of operations on the symbolic of operations on symbolic structures, while preserving the coherence and the accuracy of the deductions

- <u>Deductive reasoning</u>: Valid and rigorous, the deduction of valid knowledge is done by applying by applying general knowledge to particular cases. It is therefore the verification of theories based on patterns such as the "Modus Ponens"
- <u>Inductive reasoning</u>: Inferring laws and developing theories, from the abstraction of experimental facts. It is thus the passage to the general case or the development of a theory, this reasoning is called plausible.
- <u>Adductive reasoning</u>: Plausible, it seeks to attach causes to the premises, makes hypotheses from the observation of particular facts.
- <u>Approximate reasoning</u>: From approximate hypotheses, it is necessary to mechanisms of approximate reasoning, efficient and able to take into account the imperfections of this type of to take into account the imperfections of this type (Nes, 2013) of knowledge, to do this, it is necessary to :

Define a representation of uncertainty and imprecision.

- Extend the reasoning scheme, to take into account new aspects.

-Analogical reasoning: in humans, analogy is a very natural cognitive reasoning in various natural cognitive reasoning in various domains such as the understanding of natural language, or the analysis of visual scenes. The reasoning by analogy consists in putting in correspondence two situations in a given situations in a given universe, and deducing the solution of a new situation according to

Known solutions of situations already encountered.

The main steps of analogical reasoning (from cases) are :

- **Recall**: selection of a situation deemed similar to the new situation.
- **Adaptation**: solving the new situation, based on the similar situation selected.
- **Memorization**: learning, validation of the new situation, and its eventual Memorization [9]

3.2.7Advantages of logical representation:

- Logical representation enables us to do logical reasoning.
- Logical representation is the basis for the programming languages.[9]

3.2.8 Disadvantages of logical Representation:

- Logical representations have some restrictions and are challenging to work with.

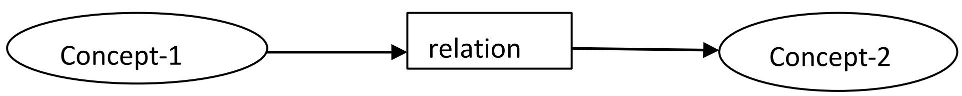
- Logical representation technique may not be very natural, and inference may not be so efficient.[9] ***3.3 Conceptual graphs:***

3.3.1 Definition:

This mode of knowledge representation has been defined by Sowa (1984)

The CG is meant to be an intermediate language between formalisms intended for automatic processing and natural language. The graphical (Nes, 2013) (Pol,2007) representation is a finite, oriented, connected, and labeled graph which has two types of nodes (bipartite)

- Conceptual node: a rectangle represents an entity;
- Relational node: an ellipse represents a property or a relationship; CG are based upon the following general form:



This may be read as: "*The* relation *of a* Concept-1 *is a* Concept-2". The direction of the arrows assists the direction of the reading. If the arrows were pointing the other way, then the reading would be the same except that Concept-1 and Concept-2 would exchange places (i.e. "*The* relation *of* Concept-2 *is a* Concept-1").

As an alternative to the above 'display' form, the graphs may be written in the

They can be noted in graphic or textual form (also called linear form).

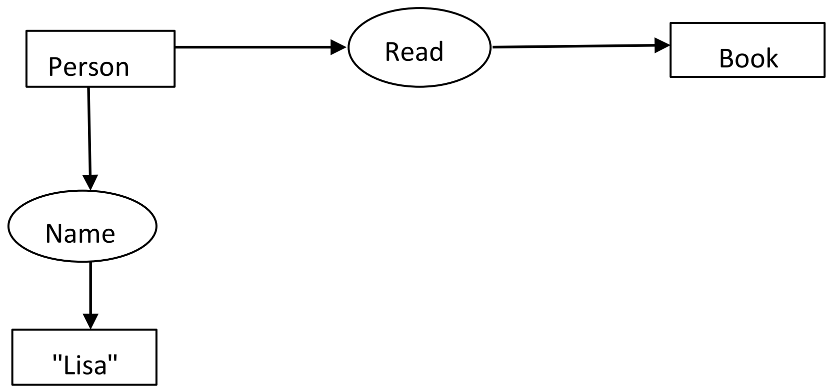
Following 'linear' text-based form:

[Concept_1] -> (relation) -> [Concept_2].

The full stop '.' signals the end of a particular graph. Consider the following Example:

Lisa reads a book.

- **Graphical representation:**



- **Textual representation:**

[« lisa »] <- (name) <- [person] -> (reads) -> [book].

In CG, type labels belong to a type hierarchy. Thereby:
Manager < Employee. ("A manager is an employee")
This means Manager is a more specialized type of the type Employee i.e. Manager is a *subtype* of Employee. Alternatively, this can be stated as Employee is a *supertype* of Manager. (Subtypes and supertypes are analogous to subclasses and superclasses in objectorientation, thus a subtype inherits the characteristics of its supertypes.) Similarly, the remainder of the hierarchy may be:

Employee < Person.(Pol, 2007)
Person < living being. living
being < Entity.
Entity < T.
Funding-Request < Request.
Request < Act.
Act < Event.
Act < T.

Sowa provides a conceptual catalog that includes a representative set of hierarchical concepts, as well as relations .It also shows the context in which a type label should be used. For example an Act is an Event with an Animate agent:

[Act] -> (agent) -> [Animate].

The type denoted as 'T' means the universal supertype. It has no supertypes and is therefore the most general type.

A subtype can have more than one immediate supertype

- **Simplification**

When two or more conceptual relations of a GC are identical, i.e. having the same type, same arity, linking the same types of concepts, then we proceed to the simplification of the relations.

having the same type, same arity, linking the same types of concepts, then we proceed to . Simplification, which consists in removing redundant information(Nes, 2013)

- **Restriction**

This type of operation concerns the restriction of concepts and the restriction of relations, in both cases, the restriction is the reasoning that specializes generic types of concepts/relationships by more specialized types of concepts/relationships.(Nes, 2013)

- **The Join**

Given a conceptual graph "C", having C1,C2,...,Cn concepts identical to the concepts D1,D2,...Dn of another conceptual graph " D ". The join operation of the two GCs "C" and "D" produce a conceptual graph "R", which is obtained by removing the concepts D1,D2,...Dn and by connecting to C1,C2,...,Cn the arcs which were connected to the nodes D1,D2,...Dn in thegraph "D". (Nes, 2013)

- **The Projection**

This operation is a morphism of graphs; it is the basis of the search mechanism in conceptual graphs.

Search mechanism in conceptual graphs allows the calculation of the specialization relation.

The search for a projection of a graph "H" in a graph "G" is seen as the search for the "inclusion" of the information represented by "H" in "G". If this is the case, we say that " G " is a specialization of " H " or " H " subsumes " G
(Nes, 2013)

3.4 Description logics:

At their origins, the description logics "LDs" were intended to provide a formal semantics for semantics to semantic networks and Minsky frames, so that these models could be to provide these models with tools for reasoning. It is a family of representation formalisms (KR), in an application domain based on the KL-One language of Brachman, which itself Brachman's KL-One language, which itself comes from semantic networks and frames.
The main constructors of description logics are :
Concept: expression used to denote sets of individuals.
Role: Denotes binary relations between individuals. In the knowledge base, we can clearly distinguish intentional knowledge which is knowledge describing the terminology of the domain, and extensional knowledge that specifies a given problem .
Following this analogy, a knowledge base presented by the description logics, is formed by the couple (T-Box, A-Box), such as: " Figure 2.2" and " Figure 2.3”

- **TBox:** it is the terminology, or the vocabulary that describes the concepts and the roles of the application domain, this component specifies the structural knowledge, it is the intentional knowledge.

In addition to atomic definitions of concepts and roles, DLs allow to build more complex Construct more complex definitions using a set of concept constructors (negation, conjunction, disjunction, restriction
(negation, conjunction, disjunction, restriction etc.). For example we define the concept "Woman" as being a "person" of "female" sex, by the statement:
Woman ≡ Person ∩ Female.

- **ABox**: also called assertional, contains individuals are assertions about concepts and relations, with certain precautions.(Nes, 2013)

and on the relations, with certain precautions for the relations which are not primitive, these assertions are the factual or extensional knowledge of the domainFemale ∩ Person(Amina) is a LDs formul

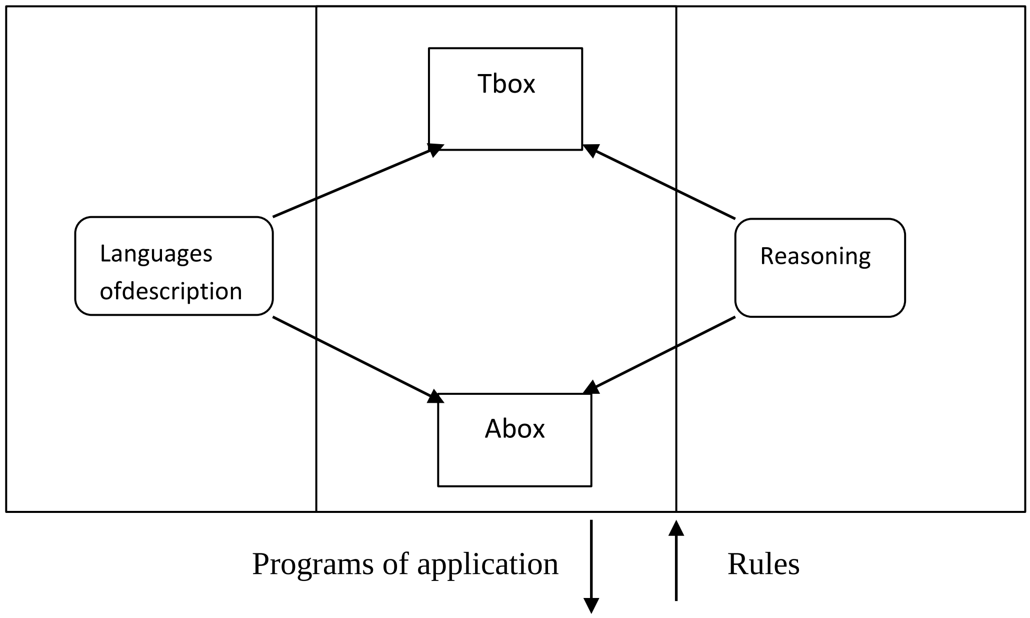


Figure 2.2: Architecture of knowledge bases in LDs**(Baa, 2003)**

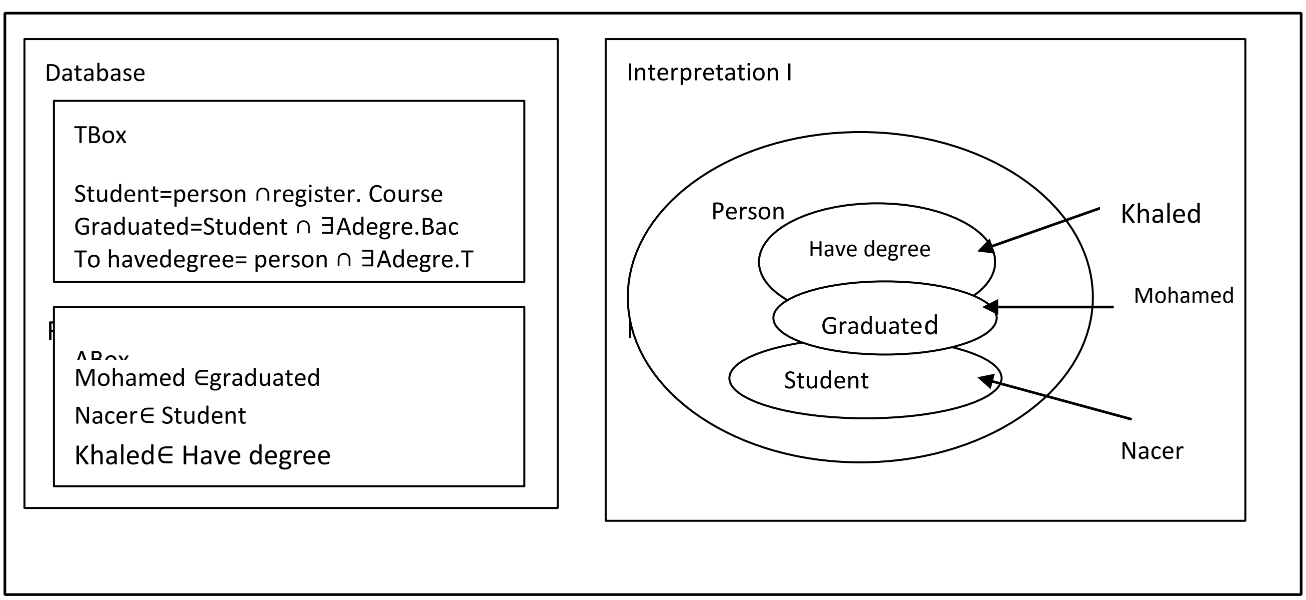


Figure 2.3: Knowledge base and interpretations in DLs(Nes, 2013)

3.5 Ontology:

Ontology is a branch of metaphysics that deals with the nature of beings, as being, it captures and structures knowledge in a way that allows it to be shared and reused. In artificial intelligence, ontology means the computer artifact that allows representing and manipulating the knowledge of knowledge of a domain by giving its components a semantic while specifying their relations. Their relationships.(Nes, 2013)

3.5.1 Goals of ontology use:

Several aspects guide the use of ontology's in AI applications:

- Communication: to enable unambiguous communication between humans and/or organizations, as opposed to organizations, as opposed to natural (Nes, 2013)language, the terms defined in the ontology have their own their own semantics.
- Interoperability: the ontology serves as an intermediate model for translation between models of different collections of objects, it is an exchange format.
- Systems engineering: the ontology can assist the process of building a system specification, model the process documents and avoid ambiguity in the specification

3.5.2 Basic elements constituting the ontology:

The basic elements of the ontology are :

a. Object: it is a representation of the real object, from the ontology point of view it is the individual.

b. Property: a property of an object or a collection of objects is a quality, a characteristic of the object, its existence is characteristic of the object, its existence(Nes, 2013) is inherent to the object.

c. Predicate: is an assertion about the domain abstraction, accepted as true, and expresses properties and constraints.

d. Class: and abstraction of a collection of objects with common properties. From point of view of ontology's it is a definition of a set of objects with their attributes.

e. Term: is a simple name, expression, formula or symbol that designates the object.

f. Intension: the intension is the set of properties that characterizes a concept.

g. Extension: Denotes a set of objects that correspond to a concept.

h. Concept: A concept is the unit of knowledge; it is characterized by the notion of intentionality, the concept represents an object that is expressed by a term, The concepts are organized in taxonomy by relations of subsumption, the extension of a The extension of a concept is the set of its instances, while its intension is the set of its attributes, properties and constraints.

i. Relation/Function: The relations are the set of associations and interactions between concepts that concepts that allow the construction of complex representations of the knowledge of the domain. These relations allow to structure hierarchically the concepts of the domain include the associations of: generalization - specialization (subclass), part of (Aggregation), etc. Functions are special cases of the relations

Conclusion:

In this chapter, we have reviewed the main concepts dealing directly with Knowledge representation with the associated reasoning mechanisms.

We also discussed the model of knowledge representations and how each model works, and its advantages and disadvantages.

The next chapter will present some measures of similarity in hierarchy.

Chapter 03:

Similarity Measurement Approaches

Introduction:

Humans use similarity for storing and retrieving information, to compare new situations to similar experiences in the past; also category learning and concept formation hinges crucially on similarity, While computers can process the binary decision of equivalence or non-equivalence of two things very fast and precisely, it poses a complex and non-trivial problem to compute similarity.

2.Use of similarity measures:

Measuring semantic similarity between concepts is an important problem in web mining and text mining which needs semantic content matching.

- Many real-world applications make use of similarity measures to see how two objects are related together. We can use these measures in the applications involving Computer vision and Natural Language Processing, for example, to find and map similar documents. One important use case here for the business would be to match resumes with the Job Description saving a considerable amount of time for the recruiter. Another important use case would be to segment different customers for marketing campaigns using the K Means Clustering algorithm which also uses similarity measures.(Khe, 2019)
- Similarity is at the heart of several works in different fields such as data analysis, case-based reasoning, pattern recognition, problem solving, learning, transfer, ..(Nes, 2013)
- The semantic similarity which expresses a connection between two concepts is a capacity abstract from man, machines do not have the ability to interpret it. It is obvious for a human that the concepts "pen" and "paper" are much more related than "temperature" and" chair ". This state of affairs, which is difficult to formalize without recourse to semantic resources: ontology’s make it possible to show the links between concepts (hyperonymy, antonymy, etc.).(Nes, 2013)

2.1 Similarity’s in the data analysis:

Many data science techniques are based on measuring similarity and dissimilarity between objects. For example, K-Nearest-Neighbors uses similarity to classify new data objects. In

Unsupervised Learning, K-Means is a clustering method which uses Euclidean distance to compute the distance between the cluster centroids and it's assigned data points. Recommendation engines use neighborhood based collaborative filtering methods which identify an individual's neighbor based on the similarity/dissimilarity to the other users.(Mav, 2019)

2.2 Similarity in pattern recognition:

Patterns are assumed to be elements of a pattern space or hypothesis class and data provide «information» which of these patterns should be used to interpret the data. The mapping between data and patterns is constructed by an inference algorithm, in particular by a cost minimization process. Fluctuations in the data usually limit the precision that we can achieve to uniquely identify a single pattern as interpretation of the data.(Buh , 2013)

2.3Similarity in case based reasoning:

- Similarity is a core concept in case-based reasoning (CBR), because the construction of the case rule, the case Retrieval, and even adaptation to a situation, all use logic based on analogy or similarity. Whatever is here It is some confusion using similarity and similarity scales and similarity scales in CBR, in particular in domaindependent CBR systems.
- Case - based reasoning (CBR) is one of the emerging paradigms for designing intelligent systems . Retrieval of similar cases is a primary step in CBR , and the similarity measure plays a very important role in case retrieval . Sometimes CBR systems are called similarity searching systems , the most important characteristic of which is the effectiveness of the similarity measure used to quantify the degree of resemblance between a pair of cases .(War et all., 2010)

3.Semantic similarity:

3*.1 Notion of Semantic Similarity:*

- Semantic similarity, variously also called «semantic closeness /proximity/nearness» is a concept whereby a set of documents or terms within term lists are assigned a metric based on the likeness of their meaning/semantic content [10]

- So is a measure of how similar a pair of concepts is (for example event A and event B). The **similarity** between A and B is related to aspects they share in common.
- Two major notions of similarity are found in existing semantic similarity measures: commonalities and differences or the semantic distance (Figure3.1). Commonalities and differences between two representations of concepts are taken as one indicator for similarity: the more commonalities and the less differences, the higher is the similarity. While some similarity assessments base their measurement on an unstructured comparison, other representations allow for a structured comparison: the common element must play an analogous role in the representation to increase the commonalties between two concepts. To apply semantic distance as a notion for semantic similarity, all concepts must be represented in a common framework with some specified metric. Some similarity measures use a multi-dimensional space as framework and the Euclidian or city-block metric for distance measurement. Semantic distance in a tree or network structure is defined by the length of the shortest path between nodes. The transformational distance is common to use for representations based on sets of transformations which can be composed and executed one after another. The distance is measured via the number of transformations or their complexity.(Sch, 2008)

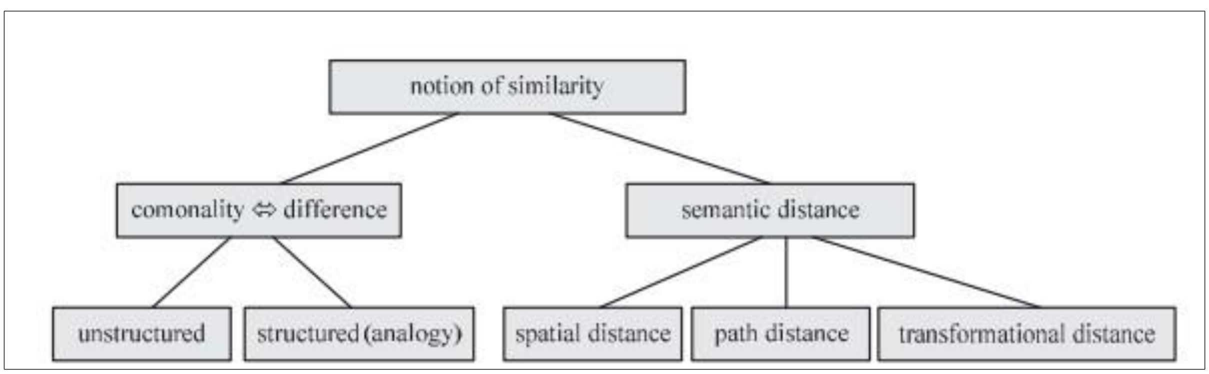


Figure 3.1; Different notions of similarity**(Sch, 2008)**

3.2 Semantic similarity measurement:

As a fundamental concept in theories of perception, behavior, social bonding, learning, and judgment, the notion of similarity has been extensively studied for several decades. In cognitive psychology, similarity is defined as the degree of resemblance between two perceptual or conceptual objects. Semantic similarity reflects the relationship between the meaning of two concepts, entities, terms, sentences or documents. Depending on the structure of the application context and its knowledge representation model, various similarity measures have been proposed.

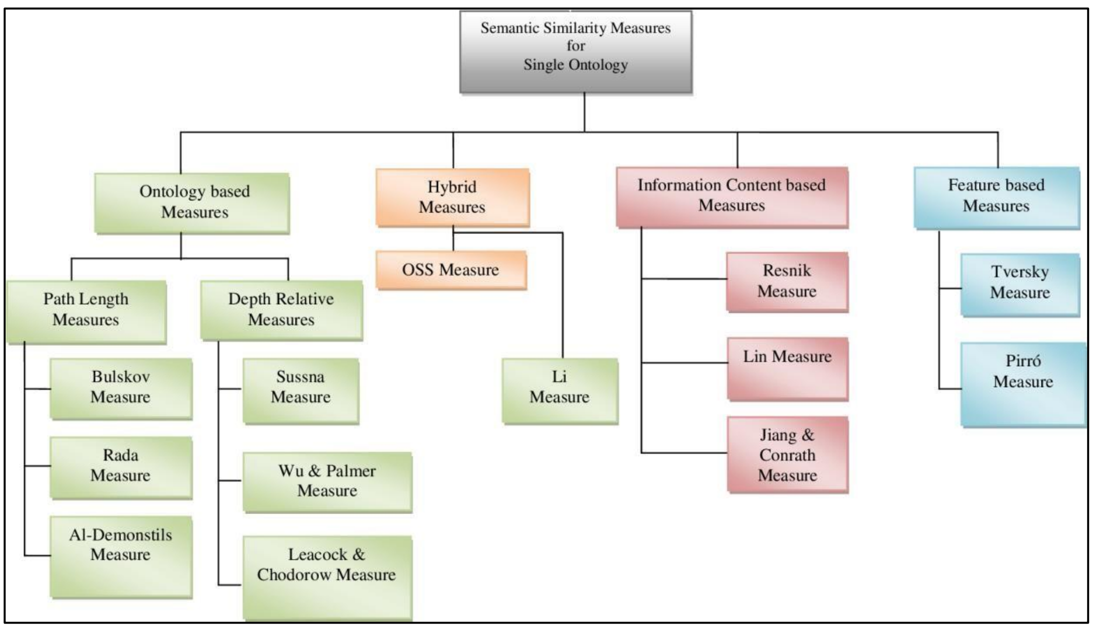


Figure3.2: categories of Semantic similarity measures [11]

Generally the result obtained from hyponym/hypernym relation is regard as similarity between concepts(Lin et all, 2013)

4.1 Ontology based measures:

4.1.1Depth relative measures:

4.1.1.1 Wu & Palmer measurement:

- Wu and Palmer introduced a scaled measure (WuP, 1994). This similarity measure takes the position of concepts c1 and c2 in the taxonomy relatively to the position of the most specific common concept lso(c1,c2) into account. It assumes that the similarity between two concepts is the function of path length and depth in pathbased measures.(Lin et all, 2013)
- The similarity measure of wu&amp is based on the following principle following principle: Given an ontology W formed by a set of nodes and a root node R (Figure 3.3). Let X and Ybe two elements of the ontology whose similarity we will compute the similarity. The principle of similarity computation is based on the distances (N1 and N2) that separate nodes X and Y from nodes X and Y from the root node and the distance that separates the subsuming concept2 (CS) of X and Y from node R.(Sli et all, 2006)

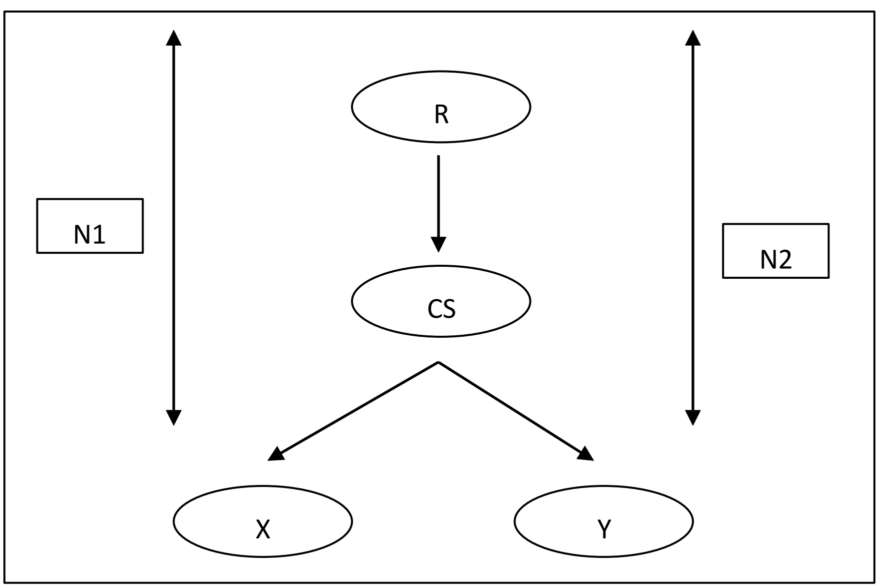


Figure 3.3: Example of an ontology extract(Sli et all, 2006)

The Wu and Palmer measure is defined by the following formula:

$$sim_{wup} = \frac{2*D(C)}{D(C_1)+D(C_2)} \qquad (1)$$

C: is the common subsumant .

C1 and C2: the concepts to compare.

D (Ci): depth from the root of the concept Ci.

4.1.1.2 Leacock and Chodorow measure:

Another method presented by Chodorow(Cho, 1998)that combines between the arc count method and the information content method. The measure proposed by Leacock and Chodorow(Cho, 1998)is based on the length of the shortest path between two Wordnet synsets. The authors restricted their attention to "is-a" hierarchical links and the path length by the global depth P of the taxonomy. The formula is defined by:

$$sim_{lc}(X,Y) = -\log\left(\frac{cd(X,Y)}{2*M}\right)(.2)$$

where:

M: is the length of the longest path that separates the root concept, from the ontology, from the lowest concept. cd (X, Y): the length of the shortest path that separates X from Y(Sli et all, 2006).

4.1.2 path length measures :

4.1.2.1 Rada, et al(1989).:

Introduces the conceptual distance measure (*c-dist*), which is calculated as the length of the shortest path between two concepts that connects the concepts through their least common subsumer (LCS). The LCS is the most specific ancestor shared by two concepts. The length is calculated by counting the number of nodes between the two concepts. (Bri et all, 2014)The *lcs-path* measure is a modification of this and is calculated as the reciprocal of the length of the shortest path as shown for concepts c_1 and c_2 in

$$Rada = \frac{1}{length(c_1,c_2)+1} \qquad (.3)$$

4.2Information Content Based method:

Another approach, in particular used in the study of semantic similarities between words in

Wordnet, improve the measures previous ones by developing the concepts with their Information Content (Information Content IC) derived from meaning labels based on the corpus or on then annotated raw corpus .**(Res, 1995)**

It was expected that each concept in WordNet has a large amount of data. The Information Content of each concept is used to calculate similarity. The more information that two concepts have in common, the more similar they are:

4.2.1Lin's Measure:

Lin proposed another method for similarity measure:

$$sim_{lin}(c_1, c_2) = \frac{2 * IC(lso(c_1,c_2))}{IC(c_1)+IC(c_2)} \qquad (4)$$

Where:lso(ci,cj): the lowest common subsumer of c1 and c2

It incorporates both the quantity of data required to state the similarities between the two concepts andthe data required to properly describe these words..(Lin et all., 2013)

As : IC(lso(c1,c2)) <=IC(c1) and IC(lso(c1,c2)) <=IC(c2), There for thevalues of this measure vary between 1 and 0.

4.2.2 Jiang's Measure:

Jiang calculated semantic distance to obtain semantic similarity Semantic similarity is the opposite of the distance

$$sim_{jiang}(c_1, c_2) = \frac{1}{Dist(c_1,c_2)} \qquad (5)$$

4.3 Hybrid methods :

The hybrid measures combine the ideas above presented. In practice many measures not only combine the ideas above, but also combine the relations, such as is-a, part-of and so on.

A typical method is proposed by Rodriguez. The similarity function includes three parts: synonyms sets, neighborhoods and features. The similarity value of the each part is assigned to a weight, and then summed together. (Lin et all, 2013)As stated before, in the paper we only concern on the hybrid measures based on is-a relation. Taking information content based measures and path based measures as parameter is commonly used. Generally one or more weight factors which can be adapted manually, are used to adapt the each part's contribution Zhou has proposed a measure, expressed by (6):

$$sim_{zhou}(c_1, c_2) = 1 - k\left(\frac{\log(len(c_1, c_2) + 1)}{\log\big(2 * (deep_{max} - 1)\big)}\right) - (1 - k) * \left(\Big(IC(c_1) + IC(c_2) - 2 * \frac{IC\big(lso(c_1, c_2)\big)}{2}\right)$$

(6)

len: the length of the shortest path from synset c1 to synset c2 in WordNet

lso: the lowest common subsumer of c1 and c2 deep_max: the max

depth(ci)

4.4 feature based measures:

Different from all the above presented measures, feature-based measure is independent on the taxonomy and the subsumers of the concepts, and attempts to exploit the properties of the ontology to obtain the similarity values. The more common(Lin et all., 2013)characteristics two concepts have and the less non-common characteristics they have, the more similar the concepts are. Features between a subclass and its super class have a larger contribution to the similarity evaluation than those in the inverse direction.

$$\text{sim}_{\text{tversky}}(c_1, c_1) = \frac{|c_1 \cap c_2|}{|c_1 \cap c_2| + k|c_1 \setminus c_2| + (k-1)|c_2 / c_1|} \quad (7)$$

Where:

k is adjustable and k∈[0,1].

4.5 statistical methods:

Statistical methods are mathematical formulas, models, and techniques that are used in statistical analysis of raw research data. The application of statistical methods extracts information from research data and provides different ways to assess the robustness of research outputs.[12]

5.Comparison and Evaluation

Category	Principle	measure	Features	advantages	disadvantages
Path based	function of path length linking the concepts and the position of the concepts in the taxonomy	Shortest path	count of edges between concepts	simple	two pairs with equal lengths of shortest path will have the same similarity
		W&P	path length to subsumer, scaled by subsumer path to root	simple	two pairs with the same lso and equal lengths of shortest path will have the same similarity
		L&C	count of edges between and log smoothing	simple	two pairs with equal lengths of shortest path will have the same similarity
		Li	non-linear function of the shortest path and depth of lso	simple	two pairs with the same lso and equal lengths of shortest path will have the same similarity
IC based	The more common information two concepts share, the more similar the concepts are.	Resnik	IC of lso	simple	two pairs with the same lso will have the same similarity
		Lin	IC of lso and the compared concepts	take the IC of compared concepts into considerate	two pairs with the same summation of IC(c1) and IC(c2) will have the same similarity
		Jiang	IC of lso and the compared concepts	take the IC of compared concepts into considerate	two pairs with the same summation of IC(c1) and IC(c2) will have the same similarity

Feature based	Concepts with more common features and less non-common features are more similar	Tversky	compare concepts' feature, such as their definitions or glosses	Take concept's feature in to considerate	Computational complexity. It can't works well when there is not a complete features set.
Hybrid method	combine multiple information sources	Zhou	combines IC and shortest path	Well distinguished different concepts pairs	parameter to be settled, turning is required. If the parameter can't be turned well it may bring deviation

Table 3.1: Comparison of Different Semantic Similarity Measures**(Lin et all, 2013)**

Conclusion:

This chapter examines different state-of-the-art semantic similarity measures based on the is-a relation in WordNet. The authors examine path-based measurements, information content-based measures, feature-based measures, and hybrid measures. We examine the principles, characteristics, benefits, and drawbacks of various measures. In addition, we give the IC metric, which is widely employed in information content-based metrics. Finally, we go over how to assess a similarity measure's performance.

Chapter04:

Proposal of a new measure

Introduction:

Identifying the similarity between concepts in an ontology is a fundamental notion, used by different techniques such as data mining, semantic information retrieval, plagiarism search, E-learning, etc.

In a domain ontology or a universal ontology such as WordNet, concepts are organized in a hierarchy linked by different types of relations such as hyponymy, hyperonymy, synonymy, meronymy and holonymy.

In this project we are mainly interested in specialization/generalization relations to perform similarity calculations in the considered hierarchy.

In chapter 3, we described the different types and categories of existing measures, we hope by this approach to improve the results given by the measures of WU-Palmer and D.Lin. In this perspective we thought of combining the two methods to increase the precision of the results and to distinguish the conceptual similarities as well as possible, this is going to result in the generation of results even more relevant and better classified for their users, it's true that the more precise we are about the similarity, the more we can better situate ourselves in relation to the objectives we are targeting.

2- Description of the approach:

Our approach uses both the notions of depth of concepts in the hierarchy used, and also the information content of the concepts to compare.

The depth-based similarity is WU-palmer, which is given by the expression:

$$sim_{wup} = \frac{2*D(C)}{D(C_1)+D(C_2)} \qquad (1)$$

C: is the common subsumant .

C1 and C2: the concepts to compare.

D (Ci): depth from the root of the concept Ci.

It is obvious that several concepts having the same depth, and common subsuming located at the same depth will have the same similarity value, however this is not always true, and can induce irregularities in the returned results.

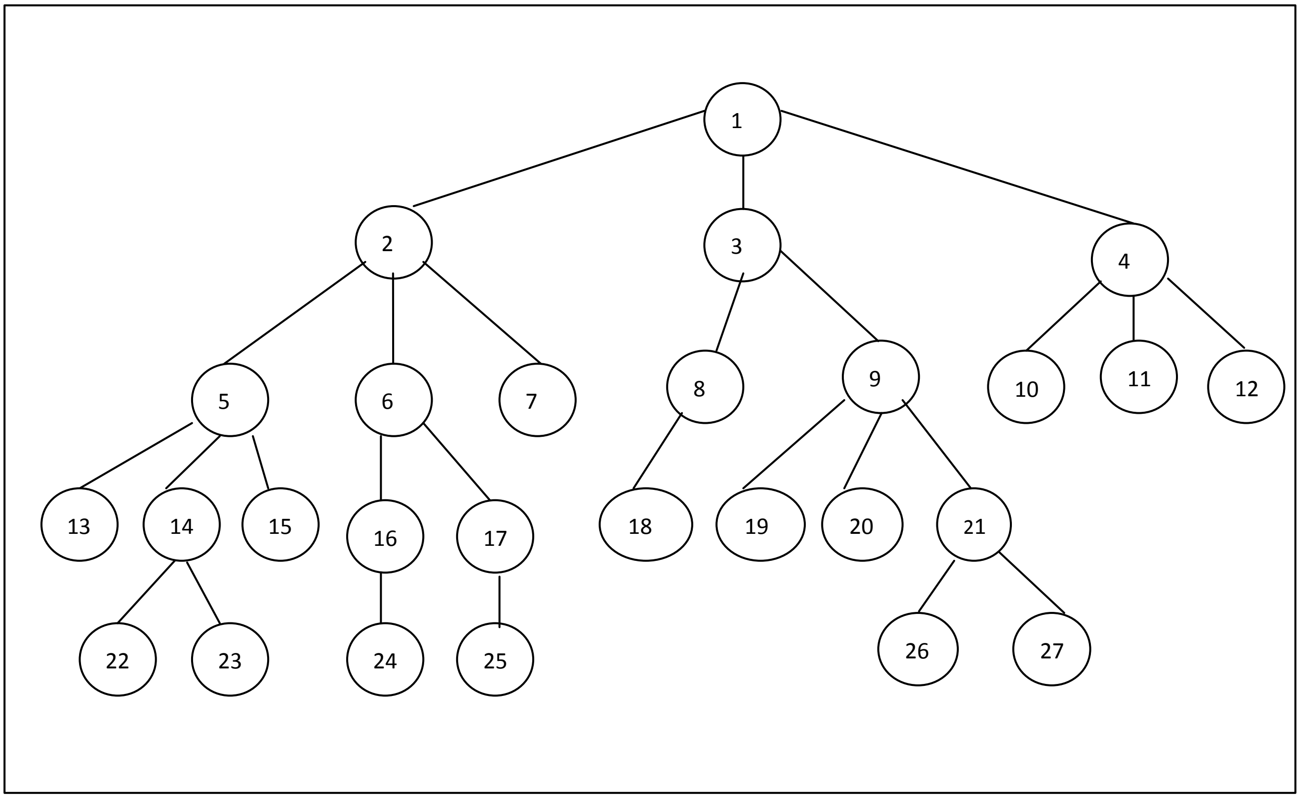


Figure4.1: Conceptual hierarchy example

We assume D(root)=1

Lcs : Least common subsumer

This hierarchy has a depth of 5

- The similarity of the two concepts '2' and '3' to their Lcs '1' will be 2/4=1/2

  It has the same value for concepts '2' and '4' in relation to their Lcs '1'. However their descendants (hyponyms) are different, the concept '4' has only leaves, so it is more specific compared to the concept '2', the semantics carried by the concepts '2' and '3' must logically be closer than for the concepts 2 and 4.
- It is the same for the concepts '13', '14' and '15' compared to their Lcs '5', normally '13' and '15' are closer, they are leaves, the concept '14' is more general.

$$sim_{\text{wp}}(13{,}15)_5 = 3 * \frac{2}{4+4} = \frac{6}{8} = \frac{3}{4}$$

$$sim_{wp}(13{,}14)_5 = 3 * \frac{2}{4+4} = \frac{6}{8} = \frac{3}{4}$$

Logically we should have : $sim_{wp}(13{,}15)_5 \geq sim_{wp}(13{,}14)_5$

- This measure generates many identical values, as an example we can take the concepts '5' and '6' compared to the Lcs '2' which gives a similarity of :

$$sim_{wp}(5{,}6)_2 = 2 * \frac{2}{3+3} = \frac{2}{3}$$

The similarity of the concepts '13' and '22' with Lcs '5' gives:

$sim_{wp}(13{,}22)_5$=3*2/(4+5)=6/9=2/3

However, the concepts '13' and '22' which are quite distant in the hierarchy, should have less similarity than for the concepts '5' and '6' which are synonyms.

The points discussed, show that this measurement lacks precision and generate few finite values.

Similarly, the similarity of D.Lin uses the information content of the concepts put in semantic comparison; the methods of evaluation of the information content are limited, the most used is that of N.Seco [2004] which is given by the expression:

$$\text{IC(C)} = 1 - \frac{\log(h(c)+1)}{\log(h)} \qquad (2)$$

IC(c): information content of concept C h(c):

the number of hyponyms of concept C h:the

total number of concepts in the hierarchy

In the above expression, we see that the information content of the concept is a logarithmic function of the number of its hyponyms, the more descendants a concept has the less informative it will be, so the information content of the root is equal to zero, and that of a leaf node approaches unity, obviously the information content in this equation is included in the interval [0,1].

And by this, D.Lin expresses the similarity based on information content of concepts by the formula:

$$Sim_{Lin}(C_1, C_2) = \frac{2 * IC(C)}{IC(C_1) + IC(C_2)} \quad (3)$$

Such as:

IC(C): the information content of their Lcs

IC(C1) and IC(C2) : information contents of the concepts C1 and C2

If we consider this formulation and apply it on the above hierarchy, we can have the following results:

- All leaf nodes of any ancestor will have the same similarity measure which is the information content of that Lcs.
-

$$Sim_{Lin}(C_1, C_2) = IC(C) \quad (4)$$

And this for any fire node C1,C2 such that C is their common subsumant.

- For all concepts having the same number of hyponyms, with subsuming concepts also having an equal number of hyponyms, we obtain the same similarity measure, regardless of their position in the hierarchy.

This is also a major drawback, as it ignores the degree of generalization/specialization of concepts in the hierarchy (see the simulation tables below)

2*.1-Formalization of our approach*

The mathematical equation used to calculate our similarity measure, referred to as 'Hs' is given by the expression:

$$\mathrm{Hs}(\mathrm{C}_1, \mathrm{C}_2) = \frac{2 * \left(\mathrm{D(C)} + \mathrm{IC(C)}\right)}{\left(\mathrm{D(C1)} + \mathrm{IC}(\mathrm{C}_1)\right) + \left(\mathrm{D(C2)} + \mathrm{IC}(\mathrm{C}_2)\right)} \quad (5)$$

In other words, to measure the similarity between the concepts C1 and C2 through their Lcs C, we considered their depths increased respectively by their informative contents.

This is an intuitive heuristic taken from the formulas of Wu-Palmer and D.Lin described below, and for which we will show some properties.

2.2-Intrinsic properties of the Hs measureThe following notations are assumed:

τ : the hierarchy

P : maximum depth of τ Ci

: concept Ci

max(τ) : the total number of concepts in τ

D(Ci) : the depth of concept Ci in τ

IC(Ci) : the information content of concept Ci

Hypo(Ci) : number of hyponyms of the concept Ci

Cs: common subsumer concept

Thus we define the measure Hs by

Hs : τ x τ →R

∀ C1,C2∈τ

Hs(C1,C2) ∈ R

$$\mathrm{Hs}(C_1, C_2) = \frac{2*(D(C)+IC(C))}{(D(C_1)+IC(C_1))+(D(C_2)+IC(C_2))} \qquad (6)$$

2.2.1- Hs is a reflexive measure:

∀ C ∈τ ,Hs(C,C)=1

Indeed:

Hs(C,C)=2*(D(C)+IC(C))/(D(C)+IC(C)+D(C)+IC(C))

=2*(D(C)+IC(C))/2*(D(C)+IC(C))=1

2.2.2- Hs is a symmetric measure:

∀ C_1,C_2∈τ

$$Hs(C_1, C_2) = \frac{2 * \left(D(C) + IC(C)\right)}{\left(D(C_1) + IC(C_1)\right) + \left(D(C_2) + IC(C_2)\right)}$$

$$= \frac{2 * (D(C) + IC(C))}{\left(D(C_2) + IC(C_2)\right) + (D(C_1) + IC(C_1))}$$

$=Hs(C_2,C_1)$

- $\forall$ **$C_1,C_2 \in \tau$ Hs(C1,C2) $\in$ [0,1]** 2.2.3

$$Hs(C_1, C_2) = \frac{2 * (D(C) + IC(C))}{(D(C_1) + IC(C_1)) + (D(C_2) + IC(C_2))}$$

We replace in this expression the informative contents of the concepts by their expression given in the equation (2) we obtain

$$Hs(C_1, C_2) = \frac{2 * \left(D(C) + 1 - \frac{\log(Hypo(C)+1)}{\log(\max(\tau))}\right)}{\left(D(C_1) + 1 - \frac{\log(Hypo(C_1)+1)}{\log(\max(\tau))}\right) + \left(D(C_2) + 1 - \frac{\log(Hypo(C_2)+1)}{\log(\max(\tau))}\right)}$$

$$= \frac{2*D(C)Log(\max(\tau)) - 2*D(C)Log(h(C)+1)}{D(C_1)Log(\max(\tau)) + D(C_2)Log(\max(\tau)) + 2Log(\max(\tau)) - \log(h(C_1)+1) - \log(h(C_2)+1)} \quad (7)$$

Lower terminal :

It is a question of proving by the expression of our similarity, given by equation (7) that its lower bound is positive and is always ≥ 0

The numerator of equation (7) is at its minimum value when :

D(C).log(h(C)+1) is at its maximum value

Two cases arise:

a) D(C) is maximal, so D(C)=P the depth of the hierarchy
   In this case, C,C1,C2 will be leaves and their
   hyponyms are null Replacing in (7) we obtain :

$$\frac{2 * PLog(\max(\tau)) - 2 * PLog(0+1)}{PLog(\max(\tau)) + PLog(\max(\tau)) + 2Log(\max(\tau)) - \log(0+1) - \log(0+1)}$$

$Hs(C_1, C_2) = \frac{P}{P+1} < 1$

b) log(h(C)+1) is max when :h(C) max(τ)-1 by
replacing in equation (7) we obtain :

$$\frac{2 * D(C)Log(\max(\tau)) - 2 * D(C)Log(\max(\tau) - 1 + 1)}{D(C_1)Log(\max(\tau)) + D(C_2)Log(\max(\tau)) + 2Log(\max(\tau)) - \log(h(C_1)+1) - \log(h(C_2)+1)}$$

$Hs(C_1,C_2) \rightarrow 0$

Similarly, the denominator of equation (7) is at its maximum value when :

Log(h(c1)+1) is min ⇒ h(c1)=0

Log(h(c2)+1) is min ⇒ h(c2)=0

Replacing in (7) we have

$$\frac{2 * D(C)Log(\max(\tau)) - 2 * D(C)Log(h(C)+1)}{D(C_1)Log(\max(\tau)) + D(C_2)Log(\max(\tau)) + 2Log(\max(\tau)) - \log(0+1) - \log(0+1)}$$

If we consider the numerator at its maximum value we put h(C)=0 to minimize

2*D(C)log(h(c)+1, and so the expression is

$$\frac{2 * D(C)Log(\max(\tau)) - 2 * D(C)Log(0+1)}{D(C_1)Log(\max(\tau)) + D(C_2)Log(\max(\tau)) + 2Log(\max(\tau)) - \log(0+1) - \log(0+1)}$$

$Hs(C_1,C_2) = \frac{2 * D(C)}{D(C_1)+D(C_2)+2} < 1$

Upper terminal:

It is also a question of showing that the upper bound of the measure Hs, does not exceed the unit, for that we also have two possibilities, that the numerator is at its maximum value and/or that the denominator is at its minimum value. the numerator of expression (7) is at its maximum value when : 2*D(C)log(h(c)+1) is minimal, so we conclude that h(C)=0 and The denominator is at its minimal value, i.e. :

Log(h(C1)+1) is maximal which gives h(C1) max(τ)-1

Log((h(C2)+1) is maximal which gives h(C2) max(τ)-1

By replacing in (7) we deduce:

$$= \frac{2 * D(C)Log(max(\tau))}{D(C_1)Log(max(\tau)) + D(C_2)Log(max(\tau))}$$

$= \frac{2*D(C)}{D(C_1)+D(C_2)}$we find the Wu-Palmer equation

3- Implementation:

3.1- Experiment in the wordnet hierarchy:

The sub-hierarchy below is taken from the WordNet taxonomy, it is relative to one meaning out of four of the word "red" chosen as a noun (there are in fact 4 synsets relative to the synonyms of the word "red")

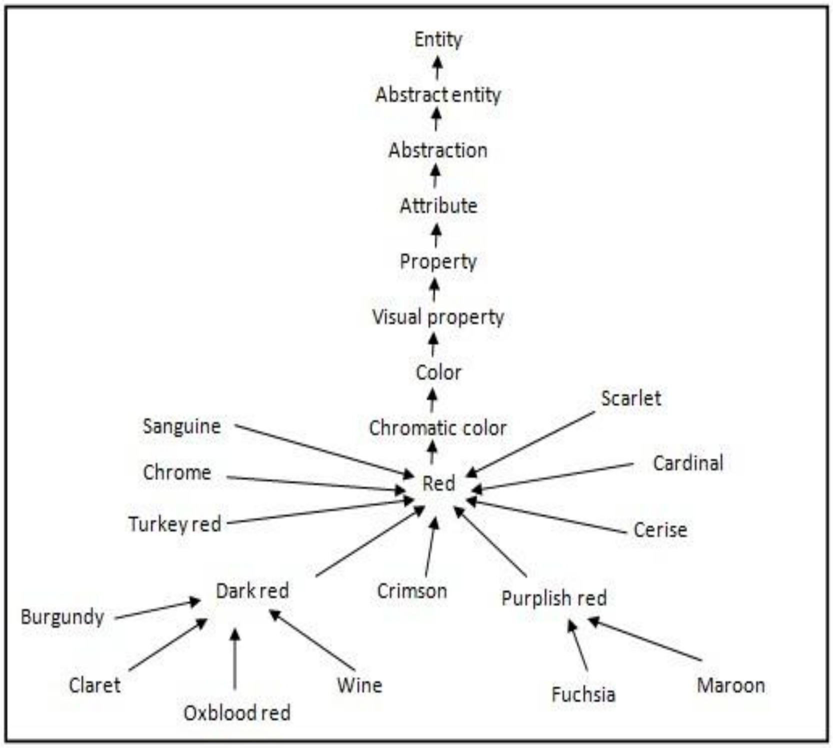


Figure4.2: Sub-hierarchy for the first meaning of the word "red" in WordNet

The application of our measure on this sub-hierarchy gives the results grouped in table 1

(c1,c2)	*d(c1)*	*h(c1)*	*d(c2)*	*h(c2)*	*d(c)*	*H(c)*	*Ic(c)*	*Ic(c1)*	*Ic(c2)*	*Hs(c1,c2)*
claret,wine	*11*	*0*	*11*	*0*	*10*	*4*	*0,60*	*1*	*1*	*0,8839*
oxblood,maroon	*11*	*0*	*11*	*0*	*9*	*15*	*0,32*	*1*	*1*	*0,7769*
darkred,maroon	*10*	*4*	*11*	*0*	*9*	*15*	*0,32*	*0,60*	*1*	*0,8248*
red chromatic color	*8*	*16*	*7*	*17*	*6*	*18*	*0,28*	*0,30*	*0,29*	*0,8051*
wine, color	*7*	*17*	*11*	*0*	*6*	*18*	*0,28*	*0,29*	*1*	*0,6511*
crimson,cerise	*10*	*0*	*10*	*0*	*9*	*15*	*0,32*	*1*	*1*	*0,8475*
color, abstraction	*3*	*21*	*7*	*17*	*2*	*22*	*0,23*	*0,24*	*0,29*	*0,4240*

Table 4.1: Hs similarity applied to the hierarchy in Figure4.1

After explaining and demonstrating our proposed measure theoretically and formerly, we are all ready to implement it programmatically to do so we need to use some tools and libraries mentioned below.

3.2-Used languages and libraries:

- **Python:** It is the most common open source programming language used by computer scientists. Among other things, Python allows developers to focus on what they are doing, not how they do it. It frees developers from the formal constraints that plagued their days with older languages. We have use that language to create our web application.
- **Javascript:** is a programming language that allows you to implement complex mechanisms on web pages. Anytime a web page displays more than just static content - content that updates at scheduled times, interactive maps…etc. We have use that language on browser side calculation to get good performance.
- **Treant.js:** a javascript library that allow us to draw and display complex trees data structures easily on browser pages.
- **Wordnet**: Is a large English word vocabulary database. Nouns, verbs, adjectives, and adverbs are grouped into sets of cognitive synonyms called "synonyms," each of which expresses a specific concept. Synsets are linked by conceptual semantics and lexical relationships such as hyponyms and antonyms. We used this database in our project to get the latest conceptual information and data.
- **NLTK**: is one of leading platform used to build natural language application with python, it provides easy-to-use interfaces to over 50 corpora and lexical resources such as WordNet and that's why we chose to use it.
- **Flask**: is a small, lightweight Python web framework, that provides useful tools and features that make creation web application easy process, we have used that library to create our server.

3.3-Hierarchy generation algorithm:

After talking about used libraries, we are ready to talk about the implementation part in depth, and starting from the algorithms we have implemented would be the best idea for that, these next are the algorithm we made.

As the theoretical part shows, it is clear that the most important part of our project is the Hierarchy, without that tree we can't calculate any relation between the concepts, so the first struggle here was to propose a valid data structure that can represent complex Hierarchy efficiently, to do so we have proposed a data structure that follows these rules:

- Hierarchy has a single root that item represents the top item of our tree, and that item has a value of depth equal to zero.
- Every node represents a concept of the Hierarchy.
- Every node has two attributes, the first one is the name of the concept which we got from the WorldNet synsets, and the second is an array of children of that concept.
- Every item of the child array represents a node that follows the same rules above.

following these rules will make hierarchy generation an easy process, now we are going to talk about the function that creates hierarchy, to explain the base idea of that function we are going to list these steps, while the implementation on (figure 4.3):

- Extract the concept synset using the WorldNet synset function.
- Create a root node containing the name of the concept, and list of children nodes by calling another function used to extract children's, while saving important data such us node depth and number of descendants that will be used on the calculation process.
- Repeating the same step for each children, until there is no longer children left for concepts.
- After completing the hierarchy, we only need to append the parents of the root node and shift the nodes so the top node will take the root position and now we have our hierarchy.

```
DefgenerateConceptTree(word):
concept = wn.synset(f'{word}.n.01')concept_name =
getConceptName(concept)tree =
{"text":{"name":concept_name},
"children":conceptChildrenGetter(concept,0)}appendConceptParents(
concept,tree)return(tree)
```

Figure4.3: Hierarchy generator function

And this the function that generates fills the children array :

```
DefconceptChildrenGetter(concept,depth=0,parent=None):result
= []childrens = concept.hyponyms()conceptName =
getConceptName(concept)iflen(childrens)==0:
return[]result = [{
"children":conceptChildrenGetter(child,depth+1,conceptName),
"text":{"name":getConceptName(child)}}forchildinchildrens
]returnresult
```

Figure4.4: Children array completion function

And last and not least this the function used after completing the hierarchy,toappend the parents of the root node and shift the nodes so the top node will take the root position and now we have our hierarchy.

```
DefappendConceptParents(concept,tree):
 hyp = lambdas:s.hypernyms()parents =
concept.tree(hyp)whilelen(parents)==2:
 parent  =parents[1][0]
parents = parents[1]
 tree = {"name":getConceptName(parent),"children":tree}returntree
```

Figure4.5: Append hierarchy concept root hypernyms (parents)

3.4. Our project work flow:

After completing the base algorithm of our project now we are ready to make an interface to use these algorithms, that was a hard step since there is a lot of options that we can use (python tkinter, Qt) these are some of the libraries used to create desktop application with python, but we couldn't chose these for a valid reason we couldn't find a python library that can draw our hierarchy, after a long research for an alternative solution we have found the TreantJs library which was the perfect drawing tool that can fix our problems, but since it's a javascript library that's mean we need a html page to use it, so for that we have decided to make our interface using web application where we have used flask python lightweight framework to implement that.

We can explain our application on these main steps:

- An HTML page that reads a word from the user.
- Send the written word to our flask application server.
- Generate the corresponding hierarchy of that word and send it back to html page, moreover send the information of each concept.
- Use the hierarchy on our web page with TreantJS to draw it in the page.
- Reads chosen two concepts from the user and calculate the similarity of these concepts directly in Javascript.

the main question you can ask yourself here is why we have chosen to calculate the similarity with JavaScript, that's easy doing that on JavaScript will save a lot of time since it is so fast, and that will reduce the pressure on our flask python server.

These are some of the screenshots from our application:

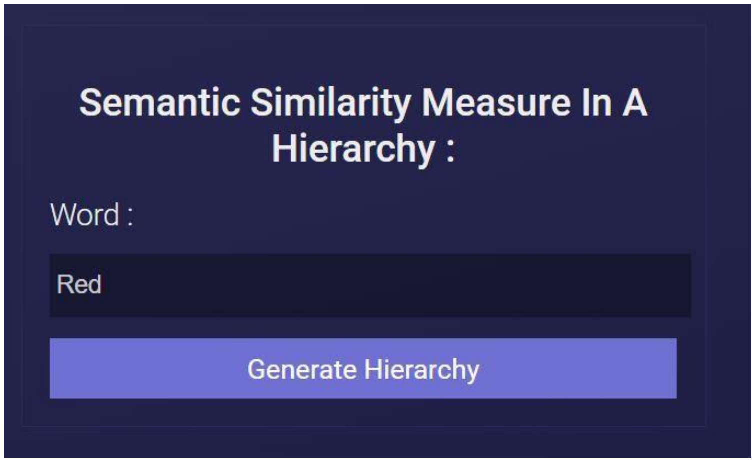


Figure4.6: Web application home page

As the figure 4.6 above shows we have a word input to read from users.

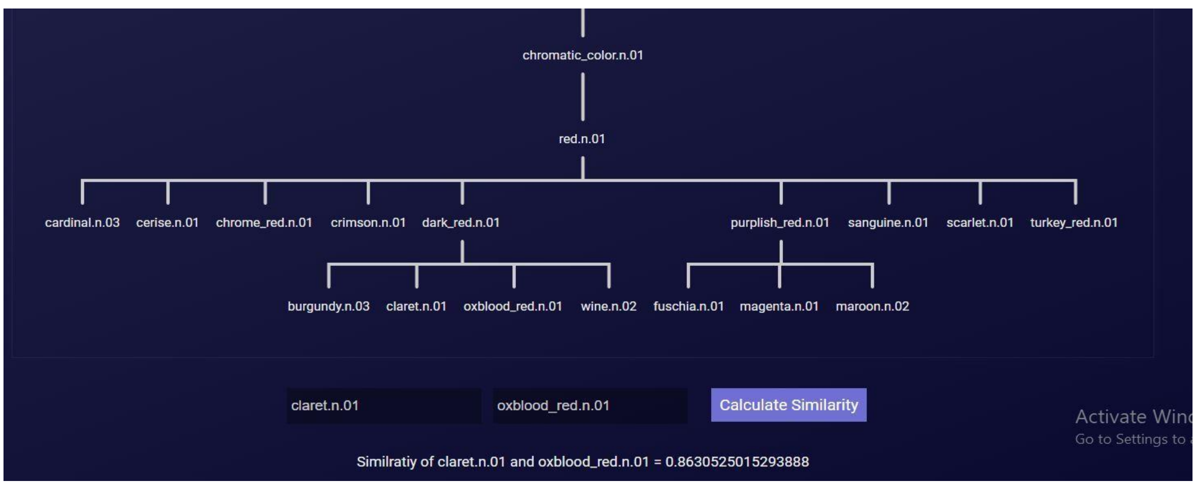


Figure4.7: Web application result page

Where the figure 4.7, shows the destination page of our application, in that page we draw the Hierarchy and give the user two inputs to choose concepts, than calculating the result and showing them on the same page very quickly.

As we said before we have used javascript to calculate your proposed similarity measurement, doing that will save a lot of time in the calculation process, and even more it will make our Flask application faster, and these are the Javascript implementation of our measurement:

```
 FunctionIC(concept){
 // Number HyponymeslethC =
treeData[concept].child
 return (
1- ( Math.log(1+hC)/Math.log(totalConceptNumber) )
 )
}
```

Figure4.8: Information ContentImplementation

In figure 4.8 we can see the way we have implemented the Information Content law using the data we got from the Python application, and for the next figure 4.10 we can see the way we have implemented our proposed solution, whereas for the case of concept1 is the same of concept2 we only return 1, otherwise, we will get the common parent between the two concepts, the value of IC and depth for each of the concepts and the shared pattern then we just apply our measurement.

```
Function    SimilarityCalculation(concept1,concept2){if(concept1==concept2)
return1;letcommonParent = getCommonParent(concept1,concept2)
letConcept1IC = IC(concept1)

letConcept1Depth = treeData[concept1].depthletConcept2IC = IC(concept2)
letConcept2Depth = treeData[concept2].depthletCommonParentIC =
IC(commonParent)     letCommonParentDepth =
treeData[commonParent].depthreturn (
        (2*(CommonParentDepth+CommonParentIC))/(
(Concept1Depth+Concept1IC) + (Concept2Depth+Concept2IC) )
    )

}
```

Figure4.9: Our proposed measurement implementation

3.5 - experiments in a generic hierarchy

In the tables: table 4.2 we have taken some data, in all 40 entries with distinct depths and informative contents on a hierarchy of experimentation and this to carry out calculations of our similarity and to compare them with those which one would obtain by the measures of Wu-P and D. Lin

N°	D(C)	H(C)	IC(C)	D(C1)	H(C1)	IC(C1)	D(C2)	H(C2)	IC(C2)	hr	Wu-P	D.Lin
1	1	18	0,2809	12	2	0,7317	10	2	0,7317	0,1092	0,0909	0,3838
2	1	14	0,3386	12	4	0,6069	9	1	0,8307	0,1193	0,0952	0,4710
3	2	12	0,3735	14	3	0,6614	8	1	0,8307	0,2021	0,1818	0,5007
4	2	12	0,3735	14	3	0,6614	15	1	0,8307	0,1557	0,1379	0,5007
5	2	11	0,3931	12	5	0,5624	4	1	0,8307	0,2752	0,2500	0,5643
6	2	11	0,3931	3	2	0,7317	3	5	0,5624	0,6562	0,6667	0,6075
7	3	5	0,5624	4	1	0,8307	5	0	1,0000	0,6578	0,6667	0,6144
8	9	8	0,4634	11	9	0,4376	11	4	0,6069	0,8213	0,8182	0,8872
9	7	15	0,3228	10	2	0,7317	12	1	0,8307	0,6216	0,6364	0,4132
10	2	15	0,3228	5	2	0,7317	5	1	0,8307	0,4018	0,4000	0,4132
11	5	3	0,6614	5	8	0,4634	6	3	0,6614	0,9339	0,9091	1,1761
12	2	4	0,6069	3	6	0,5247	4	0	1,0000	0,6116	0,5714	0,7961
13	2	11	0,3931	3	6	0,5247	4	0	1,0000	0,5614	0,5714	0,5156
14	2	18	0,2809	3	6	0,5247	4	2	0,7317	0,5525	0,5714	0,4471
15	10	6	0,5247	12	4	0,6069	15	6	0,5247	0,7482	0,7407	0,9274
16	4	2	0,7317	5	0	1,0000	5	0	1,0000	0,7886	0,8000	0,7317
17	4	4	0,6069	5	0	1,0000	5	0	1,0000	0,7678	0,8000	0,6069
18	3	3	0,6614	4	0	1,0000	5	0	1,0000	0,6657	0,6667	0,6614
19	3	5	0,5624	4	5	0,5624	4	5	0,5624	0,7808	0,7500	1,0000
20	6	3	0,6614	8	2	0,7317	9	3	0,6614	0,7243	0,7059	0,9496
21	5	3	0,6614	7	4	0,6069	6	0	1,0000	0,7752	0,7692	0,8232
22	5	2	0,7317	6	1	0,8307	6	2	0,7317	0,8452	0,8333	0,9366
23	6	4	0,6069	7	2	0,7317	8	3	0,6614	0,8061	0,8000	0,8713
24	6	2	0,7317	8	3	0,6614	8	0	1,0000	0,7623	0,7500	0,8808
25	14	2	0,7317	16	1	0,8307	15	1	0,8307	0,9021	0,9032	0,8808
26	6	4	0,6069	7	3	0,6614	9	4	0,6069	0,7652	0,7500	0,9570
27	10	5	0,5624	11	12	0,3735	13	9	0,4376	0,8514	0,8333	1,3866
28	9	4	0,6069	15	4	0,6069	12	3	0,6614	0,6797	0,6667	0,9570
29	9	4	0,6069	15	1	0,8307	12	0	1,0000	0,6664	0,6667	0,6630
30	12	3	0,6614	15	1	0,8307	14	4	0,6069	0,8320	0,8276	0,9202
31	14	2	0,7317	15	1	0,8307	15	2	0,7317	0,9335	0,9333	0,9366
32	14	4	0,6069	15	3	0,6614	15	1	0,8307	0,9277	0,9333	0,8135
33	10	3	0,6614	12	2	0,7317	13	1	0,8307	0,8027	0,8000	0,8467
34	8	6	0,5247	11	8	0,4634	10	3	0,6614	0,7706	0,7619	0,9331
35	7	7	0,4921	8	7	0,4921	9	6	0,5247	0,8317	0,8235	0,9679
36	8	4	0,6069	10	3	0,6614	12	4	0,6069	0,7398	0,7273	0,9570
37	5	5	0,5624	7	4	0,6069	9	5	0,5624	0,6479	0,6250	0,9619
38	9	6	0,5247	12	3	0,6614	12	2	0,7317	0,7502	0,7500	0,7533
39	6	12	0,3735	7	3	0,6614	9	5	0,5624	0,7401	0,7500	0,6105
40	5	8	0,4634	7	8	0,4634	10	8	0,4634	0,6095	0,5882	1,0000

Table 4.2: Data and measures of Hs, Wu-P and D.Lin similarities

- The first three columns (2, 3 and 4) contain the data of the common subsumption noted 'C', namely, its depth, the number of its hyponyms and the calculated value of its information content.
- The next 6 columns contain the same data for the compared concepts, namely 'C1' and 'C2'.
- The last three columns show the measured similarities, ours in red, that of Wu-P and D.lin

	Hs	Wu-P	D.Lin
Number of identical values	0	15	7

Accuracy in sample %.	100	62 ,5	82,5

Table 4.2 : Accuracy number of identical values

The graphs in the following figures show us the variations and trends of the measurements, in Figure 4.10 we have the graphs of our measurement with that of Wu-P, and in Figure 4.11, we give the graph of our measurement with that of D.Lin

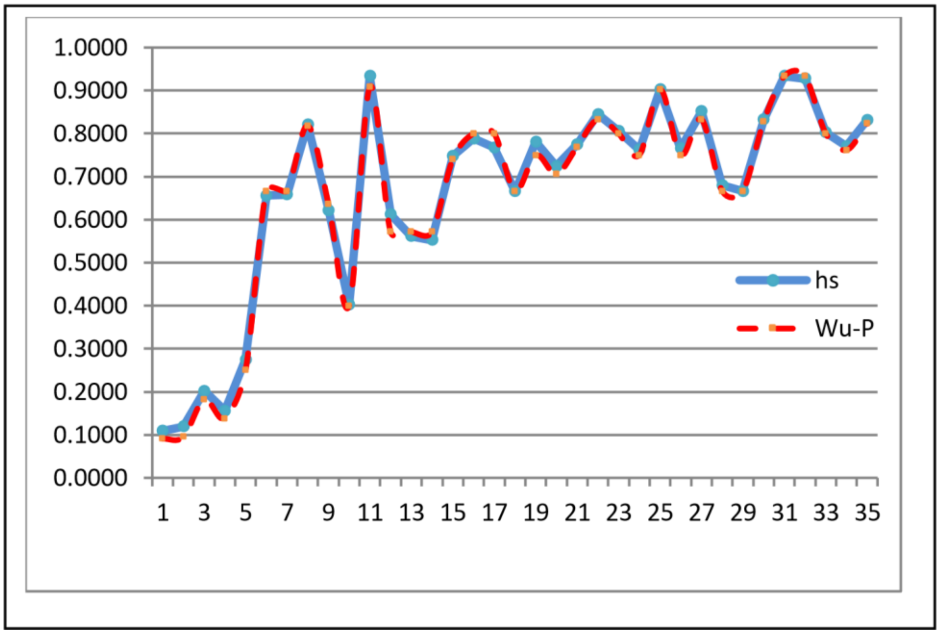


Figure4.10: Trend of Hs and Wu-P measurements

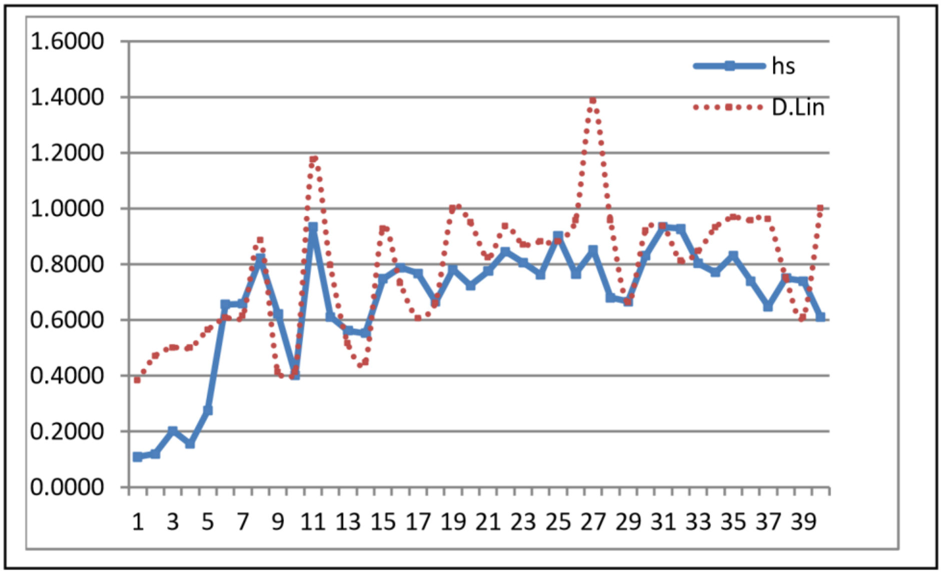


Figure4.11 : Trend of Hs and D.Lin measurements

3.6-Discussion of results:

This sample of experimental values provides us with valuable information regarding the improvement of Wu-P and D.Lin measurements, and their trends, indeed we can deduce from Table 4.3 above that:

Out of the 40 data entries used, we have :

- 100% accuracy for our similarity, it is less for the other two measures.
- In applications using this type of similarity measures such as for semantic disambiguation, plagiarism, and information retrieval, this could have inappropriate consequences and could lead to inaccurate and inappropriate results.

Also, seeing the graphs in Figure 4.10 and 4.11 we see that:

- There is a harmony in the variations between our 'Hs' measure and the 'Wu-P' measure, because the curves are almost superimposed, with the difference that for the 'Wu-P' measure we have a restricted distribution.
- A better precision for our 'Hs' measurement.
- There are fluctuations and concordances between our 'Hs' measurement and that of 'D.Lin', this is due to the use of depths that affect the Lin measurement. It should be noted that in some cases, the measurement of 'D.Lin' sometimes exceeds the unit

Conclusion:

In conclusion for this chapter, we have explained briefly some of the main and famous semantic measurements used in a hierarchy of concepts, which are depth-based similarity also known as WU-palmer, and the Information Content (IC), explained the concept of this measurement mathematically.

After that, we have presented our proposed similarity measurement that combines the power of depth-similarity of WU-palmer law, and the Information Content, then we have proved our proposed similarity In theory and practice, in the end, we have to make a web application that generates fully custom hierarchies for users, and help to calculate the semantic similarity of concepts and that was our goal

GeneralConclusion:

The main goal of this work is to propose a new inter-document similarity measure that takes into account the contents of documents and not only the syntactic form.

We have started our work by explaining mathematically some of the most used semantic measurements used for hierarchal concepts, which are depth-based similarity also known as WU-palmer, and the Information Content (IC), done some examples using these two measurements.

After that, we have presented and proved modeling and experimentation of a new measure of semantic similarity in a conceptual hierarchy that uses the WordNet taxonomy dataset, this measure can evaluate the similarities between two concepts, the proposed measure is based on two measures WU-palmer and Information Content.

We compared our measure with the WU-Palmer measure and IC, we find that our similarity measure has an accuracy of 100% which is better than the other measurements.

After being sure of the results of our model theoretically, we started the implementation using the Python language, where we have created an algorithm that generates conceptual hierarchy using the WordNet database and ended that by creating the interface that uses these algorithms, the results were quite perfect and compatible with our start hypotheses.

However, there is always room for improvements, and from our perspective, huge research can be done, for example making our semantic measure works with sentences not just words, doing this will open a door to huge improvements for current semantic ontology's.

This work is the subject of submission to the conference" the 7th Smart city application Castelo Branco Portugal "

****18-21 October 2022****

Bibliography:

(Baa,2003) : “the Description Logic Handbook,Theory, Implementation, Applications” [Book] Franz Baader,Cambridge University Press, 2003.

(Bir et all., 1999): “Applications of Knowledge Acquisition inExperimental Software Engineering"[Book],Birk, A., Surmann, D., and Althoff, K.-D,1999.

(Bri et all, 2014):”U-path: An undirected path-based measure of semantic similarity”[article],Bridget T. McInnes, Ted Pedersen, Ying Liu, Genevieve B. Melton, and Serguei V. Pakhomov,2014.

(Buh , 2013)**:”**Similarity-Based Pattern Analysis and Recognition ”[Book],Joachim M Buhmann, 2013.

(Cho, 1998):”Combining Local Context and WordNet Similarity for Word Sense Identification”,Chodorow, C. Leacock et M.,1998,In WordNet: An Electronic Lexical Database, C. Fellbaum, MIT Press.

(Chr,2021):"What is knowledge management and why it’s important"[Book],Iryna Chorna,2021.

(Edo, 2014)**:”**Knowledge Management Concept”[conference],Edosio, Uyoyo Zino,2014,Knowledge Management,Bradford.

(Gen et all ,2012):“Analogical Reasoning”[Book],D. Gentner, L. Smith,2012Encyclopedia of Human Behavior (Second Edition).

(Gha,2009):” Knowledge management: tools and techniques”[article],Syed RaiyanGhani ,2009, DESIDOC Journal of Library & Information Technology.

(Gup, 2022):“7 Types of Knowledge: Explicit, Implicit, Tacit, & More”[online],Disha Gupta ,2022 ,https://whatfix.com/blog/types-of-knowledge/

(kao et all,2013) : ‘’ Enrichissement de la représentation conceptuelle dans la catégorisation du texte en utilisant les mesures de similarité sémantique ‘’,Bouyacoubsoumia - Kaouadji

Sarra,Mémoire de Master en Informatique,Université Abou BakrBelkaid– Tlemcen - 2013

(kha et all ,2016):”Abstraction and Generalization in Conceptual Design Process: Involving Safety Principles in TRIZ-SDA Environment”[Book],KhairulManamiKamarudin, Keith Ridgwaya, NapsiahIsmailb,2016.

(khe,2019):”Similarity Measures — Scoring Textual Articles”[article],KherajSaif Ali,2019,Published in Towards Data Science.

(Lie, 2001):”knowledge management and its link to artificial intelligence

“[book],Liebowitz J ,2001.

(Lin et all, 2013)**:”**A Review of Semantic Similarity Measures in WordNet”[journal article],LinglingMeng, Runqing Huang and Junzhong Gu,2013, china.

(lyd et all,2017) : ‘’ Extension d’un modèle de recherche d’information pour la prise en compte

de la représentation de type wordmbedding ‘’, Melle Bouchaal Lydia. Melle LeulmiFazia,

Mémoire de Master en Informatique , Université MMTO , 2017

(Mai, 2022)**:"**Declarative & Procedural Knowledge: Differences & Uses”[online]

,Maria Airth ,2022 , https://study.com/academy/lesson/declarative-

proceduralknowledge-differences-uses.html

(Mar,2019)‘ ‘ How Important Are Semantic Networks In Artificial Intelligence ‘ ‘ [article],F.R. MARTIN, 2019, india.

(Mav, 2019): “ Calculate Similarity — the most relevant Metrics in a Nutshell “[article] Lüthe Marvin,2019.

Source spécifiée non valide.**:"** Data Information and Knowledge Management "[Online], **Mestas Juan** https://itabok.iasaglobal.org/itabok3_0-2/data-informationknowledge-management/.

(Mik et all., 2003)**:"**Managing Software Engineering Knowledge**[Book],**Rus, Mikael ,2003 .

(Nes,2013)"UN model de raisonnement pour un système de recherche semontique d information sur le web base agents**"** [these du doctorat],.NessahDjamel , 2013, algerie

(Nes et all,2016) "Towards a hybrid semantic similarity measure to set the conceptual relatedness in a hierarchy" ,DjamelNessah, Okba Kazar , Aïcha-Nabila Benharkat [article]. - 2016. , algerie

(Nür et all., 2011):"Wisdom - theblurrytopofhumancognitionintheDIKW-model?[online]**"**Andreas Nürnberger,Constanze Wenzel,Rudolf Seising,Anett Hoppe,2011, https://www.researchgate.net/figure/Data-Information-Knowledge-Pyramid-24_fig3_269081562

(Pat, 2020): **"**what is procedural knowledge and how to use it?"[online],Patel Riddhi,2020

https://www.thecloudtutorial.com/proceduralknowledge/#:~:text=Procedural%20knowledge%20can%20be%20simply,about%20abilities%20to%20do%20something.

(Pol,2007):" An Introduction to Conceptual Graphs " ,Polovina Simon,2007,Conference: Conceptual Structures: Knowledge Architectures for Smart Applications,United kingdomwooldridge mike lecture 7: propositional.

(Buh , 2013)**:"**Understand what a priori knowledge is. Learn the meaning of a posteriori knowledge and explore the difference between a priori and a posteriori"[book],Pollens Dempsey,2021 .

(Pra, 2022):"explicit knowledge: definition, examples, and methods" [online],JubinaPrabhakaran,2022 ,https://document360.com/blog/explicitknowledge/

(Raz,2020):”Knowledge Management: Guidelines and Best Practices”[online], Raza Muhammad,2020, https://www.bmc.com/blogs/knowledge-management-best-practices/.

(Res, 1995) :”Using information content to evaluate semantic similarity in a taxonomy.”,Resnik, P.,1995.

(Sağ, 2006)**:”**A new life cycle model for processing of knowledge management”,Mustafa Sağsan,2006.

(Sch, 2008)**:”**Approaches to Semantic Similarity Measurement for Geo-Spatial Data: A Survay ”[article],Schwering, Angela,2008.

(Sha):”What is a knowledge work system?”[online],HitanshSharma,https://www.quora.com/What-is-a-knowledge-worksystem.

(Sli et all,2006) :“A New Similarity Measure based”[article],T. Slimani, B. Ben Yaghlane, and K. Mellouli,2006,World Academy of Science, Engineering and Technology.

(Tiw, 2000)**:”**The Knowledge Management Toolkit”[book],Amrit Tiwana,2000.

(War et all., 2010)**:** ”Similarity measures for retrieval in case-based reasoning systems ”[book],T. Warren Liao, Zhiming Zhang &Claude R. Mount,2010.

(WuP, 1994)**:”** Verb semantics and lexical selection”[conference], Z. Wu and M. Palmer,1994, Proceedings of 32nd annual Meeting of the Association for Computational Linguistics, Las Cruces, New Mexico.

(Xav,2006) : ‘’Extraction et recherche d’information en langage naturel dans les documents

semi-structurés ‘’Xavier TANNIER – thèse de doctorat - Ecole Nationale Supérieure des Mines de Saint-Etienne –2006

[1]:https://www.getguru.com/reference/knowledge-management-process

[2]:https://cdn.citl.illinois.edu/courses/PSYCH100/chapter8/pt1-html/web_data/file19.htm

[3]:https://www.zendesk.com/service/help-center/knowledge-management-tools/#definition

[4]:https://www.kpsol.com/glossary/what-is-a-knowledge-management-system-2/

[5]:https://www.slideshare.net/SWEETLIME123/laudon-ch11

[6]:https://paginas.fe.up.pt/~als/mis10e/ch11/chpt11-2bullettext.htm

[7]:https://paginas.fe.up.pt/~als/mis10e/ch11/chpt11-4bullettext.htm

[8]:https://www.javatpoint.com/knowledge-representation-in-ai.

[9]:https://www.javatpoint.com/ai-techniques-of-knowledge-representation

[10]:https://www.igi-global.com/dictionary/information-retrieval-by-linkage-discovery/26364[11]:https://www.semanticscholar.org/paper/A-Survey-on-Semantic-Similarity-Measure-Elavarasi-Akilandeswari/7e1ece8d9041c60587c01732e60b5853bbb2f466

[12]:https://www.nature.com/subjects/statisticalmethods#:~:text=Statistical%20methods%20are%20mathematical%20formulas,the%20robustn ess%20of%20research%20outputs.