

A MIND MAP QUERY IN INFORMATION RETRIEVAL: THE 'USER QUERY IDEA' CONCEPT AND PRELIMINARY RESULTS

Rihab Ayed¹, Farah Harrathi², M. Mohsen Gammoudi² and Mahran Farhat³

¹Tunisia Polytechnic School, University of Carthage, Tunis, Tunisia
rihabb.ayed@gmx.com

²Higher Institute of Multimedia Arts of Manouba, University of Manouba, Tunisia

{harrathi.farah@gmail.com, mohamed.gammoudi@fst.rnu.tn}

³Faculty of Sciences of Tunis, University Tunis El-Manar, Tunisia
farhatmahran@gmail.com

ABSTRACT

Users in Information Retrieval are formulating since many years their queries in a bag of words that should be understandable by the system. The problem of 'bag of words' format is that it can cause the deformation of the user's information need. A question is raised in this paper to discover if there is a more faithful and richer way to users to formulate their idea of search. This paper proposes an approach for users to model their queries in Information Retrieval (IR) based on the use of a brainstorming technique (Mind Mapping). The choice of the query representation model is based on assertions concerning Human Mind and habits of thinking. In this approach, an interpretation is suggested for the use of Mind Map, based on the relative importance weight of terms. Preliminary experimentation on a Medical corpus (CLEF2009) showed the accuracy of our approach.

KEYWORDS

Information Retrieval, Query Formulation, Idea, Associations, Mind Map

1. INTRODUCTION

There are two main components manipulated by an Information Retrieval System (IRS): the user query which is expected to traduce faithfully the user information need and the documentary corpus from which the IRS selects relevant documents [1].

According to a study in 2001 [2], the average query length in the Excite search engine is 2.6 words. A Report in the beginning of 2010 [3] done by the search engine Google, asserts that the length of 54.5% of queries in Google is greater than 3 words. In a 2013-2014's period, the multi-words queries (producing a click) constitute about 45.92% of all queries [4]. In Google Mobile Search, the average English query length in 2008 is between 2.44 and 2.93 words [5]. In 2012, a study about Mobile searches [6] indicates that average English query length is 3.05 words. These statistical results could reflect the ambiguity problem of users' multi-words queries in Information Retrieval Systems.

Natarajan Meghanathan et al. (Eds) : ICCSEA, SPPR, VLSI, WiMoA, SCAI, CNSA, WeST - 2014
pp. 197-213, 2014. © CS & IT-CSCP 2014

DOI : 10.5121/csit.2014.4726

The key challenge that remains nowadays is finding ways to capture and integrate contextual information for solving the ambiguity problem of queries expression. This contextual information is important to ameliorate the assimilation of the user's multi-words query and the user information needs in general [7]. Several approaches handle the queries ambiguity by proposing three main solutions [7]: (i) user profiling and personalization (ii) query expansion (i.e. automatic expansion or by recommendation) and (iii) Relevance Feedback. These approaches are generally founding their techniques on an input: a bag of words query, to reach the user's idea of search which would lead to the output: the documents relevant to the user's idea.

Some critics [8,9] tend to say that search engines are becoming more and more intelligent, using these solutions, while the user is becoming more and more dependent and less intelligent. In fact for query formulation, users are dragging since years their unchanged way of expression: the bag of words.

The problem of bag of words expression is that it causes sometimes the deformation of a user information need. This can be due to two facts: (i) the user cannot express clearly the main words of his idea of search and the less important words (which are used to specify the context related to the idea of search) and (ii) the idea of the user is packaged in a linear form which does not express the nodes in the network of the user's mind. As an example for the second point, let Q be the query "Java course Graphical Interfaces swing". This query can be represented in the mind of the user as a related multi-nodes query and can be formulated by the sentence "A course for Graphical Interfaces in Java language using the Swing API". In the query Q, the connections between remembered words are absent, which can misrepresent the flow of the user's idea of search.

Confusions come to evolve concerning the relation between the bag of words query and the findings in Cognitive Sciences, involving questions such as: Is the bag of words form the best natural way to express the user's idea? Is it possible to find a more faithful and not philosophical way of expressing the internal user's idea? We contribute in this paper to answer these questions by studying the cognitive sciences' findings about information recall process in the human mind and comparing existing studies about user expression forms used in IR or in Cognitive Sciences. This is elaborated in order to select the best natural way for user query formulation.

In this paper, we start by presenting in the section two, the state of the art which contains: (i) a dashboard of several Cognitive Maps and (ii) researches in IR that integrate query graphs and/or weighted queries (representing levels of importance in a query). In the third section, we incorporate our contribution of query formulation by expressing user's mind. In the fourth section, the evaluation of the contribution is presented. This is followed by the outcome of this work and some motivations of further works in the section five.

2. RELATED WORK

According to [10], students nowadays are connected and surrounded by networks of information, but a few of them have developed tools of thinking which express their information in a sensed network of knowledge. In fact, in one hand, human ideas are expressed by a dominating way: the linear language. In the other hand, the mind is constantly and naturally trying to find connections between information. So, there is a gap between the *internal* networks in the human mind and the *external* linear way of human expression. The process of thinking about a domain involves reasoning through our minds' Cognitive Maps and finding a way through them [11]. Human beings use these maps unconsciously. The external modelling of these maps helps to understand a domain or a cognitive territory [11].

Following these general assertions about cognition process, in the next section, we explore the kind of maps that could replace the bag of words query.

2.1 Findings about Human Cognitive Maps

A map can be a generic term for several forms: trees, graphs, diagrams, networks, etc. In the past, Cognitive Maps defined by Tolman (1948) [12] are mental representations of physical places. Nowadays, Cognitive Maps are not defined only for physical environments but also for several mapping methods such as: *Causal Mapping*, *Concept Mapping*, etc. [12]. Another popular designation of Cognitive Maps used in Education and Learning is visual tools [13]. Visual tools are divided into three categories (see Figure 1) [14]: (i) *Brainstorming webs* which are dominated by Mind Mapping techniques (1970's), (ii) *Graphic organizers* (1980) which are very structured tools, called also task-specific graphic organizers, and (iii) The Thinking-Process Maps (shown in Figure 1) which contains two main approaches: (a) the *Conceptual Mapping* which is dominated by the Concept Mapping technique (created in 1960 and became famous in 1984) and (b) the *Thinking Maps* (1988) which combine the freedom of thinking of brainstorming tools and the structured aspect of graphic organizers.

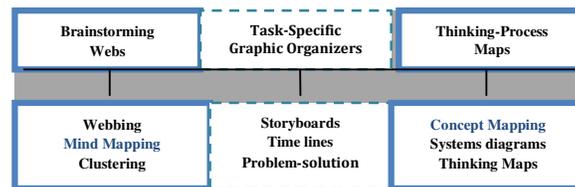


Figure 1. Clustering of Visual tools

According to our study, *Thinking Maps* [14] and *Task-specific Graphic Organizers* [10], [15] are not considered convenient for IR query formulation this is due to these reasons: (i) Graphic Organizers structure the ideas for specific goals. In fact, the use of one of the Graphic Organizers is chosen according to some characteristics relative to the user idea (For example, to choose a diagram, the user has to know if the level of his thinking is superior to three or not, or if his idea is hierarchical or not, etc.) [10], [15]. (ii) Thinking Maps are not designed for vague or fuzzy thinking process. Each Thinking Map is designed for a specific task of thinking. As an example of Thinking Maps, the *Bubble Map* [13] is used to describe an object by defining its features, using adjectives.

Further, the *Cognitive Maps* which are considered in our search as potential maps for query formulation are *Mind Maps* and *Concept Maps*. Definitions and examples of these two last maps and their techniques are mentioned respectively in [16,17,18] and in [19]. In the following section, we introduce works in Information Retrieval, representing queries as graphs (Cognitive Maps, Conceptual Graphs of Sowa, etc.).

2.2 Specifying User Context by a Query Graph

Some researchers are trying to represent the query in a graph in the phase of the indexing process (graphs are invisible to the user) such as in [20]. The goal of these works is to integrate the semantic aspect both in the indexing process and in the correspondence process between documents and queries. In this state of the art, we introduce the researches representing the query in a graph in the phase of query formulation. We are interested by these researches since they change the user behavior from formulating his need in unrelated words (bag of words), to

constructing a graph which connects the query's whole idea and specifies more the user's search context.

2.2.1 "Googling from a Concept Map"

Some studies merge to the automatic query formation based on Concept Maps. In [21], the authors propose an automatic generation of Web queries from a user's Concept Map under construction. The goal of "Googling" based on Concept Maps is to provide to users supplementary information to add to their Concept Map while they are constructing it. The steps are: (i) the construction of a Concept Map by a user, (ii) the selection of a concept from the map by the user, (iii) the map is analysed to form automatically a textual query for information retrieval. The textual query is composed of key concepts. The key concepts are selected from the map to describe its topic. The difference between an ordinary search and a Concept Map based search is that the main concept of the query is searched through the web considering its semantic context. The Concept Map-based search showed ameliorated results comparing to the Single concept-based search (in terms of similarity and coverage between the retrieved documents and the query). The inconveniences of the use of Concept Maps for query formulation are mentioned further in the section 3.1.

2.2.2 SyDoM: a Multilingual IRS Based on Knowledge Management

In order to improve the documents content representation, the system SyDoM (Multilingual Documentary System) [23] employs a model of knowledge representation named Semantic graph (an extension of Sowa's Conceptual graph)[23]. This graph is used to represent both the user query and the documents existing in a multilingual corpus. The Semantic graph is wholly constructed by a semantic thesaurus [23] of a specific domain (e.g. Vehicle Mechanics). The main goal of using Semantic graphs in this system is to retrieve concepts from documents that are related in almost the same way they could be related in the query graph. Discussion about the limits of the use of Semantic graphs in query formulation is mentioned further in the section 3.2.

2.2.3 The Google's Wonder Wheel and Knowledge Graph

In a period from 2009 until 2011, the search engine Google proposed a tool named the Wonder Wheel [24]. The tool proposed expansions of the user query, visualized by a *Mind Map*. In this Mind Map, the user query is the central node, and the associated nodes represent the related ideas to the query proposed by the system. The goal of this tool is to guess the search idea of the user (For example: the user query "Mind Maps" has an associated idea: "Mind Maps tutorial"). In this work, the Mind Map is used as a visualization tool and not a way to express the user's own associations in mind. Further, Google has beenproposing since 2012 another associative concept for searching: The *Knowledge Graph* used for retrieving associations (data or facts) related to the user query and presenting the information in an aggregated way. These associations connect information which can be entities or attributes [25]. This last Google project seems to be oriented to queries containing one-concept (For example: "*Albert Einstein*") instead of an idea of search (which can contain many associated words and evolving eventually more than one concept). Also, this project offers *Knowledge associations* in order to enrich user's network of knowledge. We are interested in this paper, by *user's mind associations* in order to improve his attempt of query formulation.

2.2.4 The Pearltrees System

The concept Pearltrees is a collaborative and social bookmarking service. It proposes to the user to formulate and to store his interests in a Mind Map[26], also it allows the user to search for related

interests of other users. In a search phase, the user can either formulate his interest in a bag of words form or select a node(pearl) from his created Mind Map(named pearl tree). When a node is selected, it is considered as well as its sub-nodes in the search. The inconvenience of the Pearl trees system is that it proposes a *semantic clustering* tool for user interests. For example, let an interest I in a form of bag of words be “*popular terms Information Retrieval*” which can be traduced by the question “*what are popular terms in the Information Retrieval domain?*”. In the Pearl trees system, the clustered nodes of I would be as follows: a super-pearl: {Information Retrieval} and a sub-pearl {popular terms}. While it appears that this clustering avoids the *user importance degree*, since the main idea of search in the interest I is {popular terms} not {Information Retrieval}.

The problem of bag of words is the non-fidelity of representation of the network in the user’s mind. In fact, there is no existence of nodes or relations. For example, the bag of words “definition concept thesaurus” could be in the mind of the user {“definition concept”, “thesaurus”}, where the node “thesaurus” is only used to specify the context of the query (“definition of a concept in a thesaurus”). So, the word “thesaurus” appears to be less important in the query than the central search idea of the user “definition concept”. A lack in a query bag of words format, is that there is no *importance degree* between used terms. Some works in literature integrate importance preferences in the query. We describe these relative works in the next section.

2.3 Adding Preferences in Queries: Weighted Query Languages

A body of research in *Fuzzy Boolean Information Retrieval*, since the 90’s [27] until 2010 [28] has followed the postulation that “weighting query terms in a different way would lead to a more effective IRS” [28]. According to authors such as [29], the query formulation subsystem in IRS has to take into consideration the imprecision and uncertainty aspects of human communication. As a potential solution, adding user weights to the query could reduce the imprecision and clarify more the user information need, by mentioning what is important and what is less important in the query. As an example of numeric weighted query languages, the works of [28],[30] propose an enrichment of the Boolean query expression as follows:

$$Q = \langle t1, w1 \rangle OR \langle t2, w2 \rangle AND NOT \langle t3, w3 \rangle \quad (1)$$

Where: $t1, t2$ and $t3$ are search terms and $w1, w2$ and $w3$ are numeric weights in the interval $[0,1]$

A critic concerning user numeric weights holds that “a human being is more capable to qualify the importance of a concept, than to quantify it” [27]. An example of the *fuzzy linguistic approach* is given as follows [27]:

$$Q = \langle t1, I \rangle OR \langle t2, FI \rangle AND \langle t3, NVI \rangle \quad (2)$$

Where: I : Important, FI : Fairly Important, NVI : Not Very Important

More query weighting methods in *Fuzzy IR* exist in [29]. In vector space IR, manual query weighting terms were present in the Microsoft Index Server. For not encumbering the user by weights, some studies tried to *automatically* weight queries by guessing terms importance such as in [31] (this study is based on a cognitive hypothesis non certified by scientists or by experimentation). Other proposals of weighting query terms differently are predicting and estimating weights based on the corpus statistics [32] or documents relevance measures [33] or on the relevance feedback mechanism [34] or on the semantic relations between query terms [35]. We are mainly interested in this paper by the user importance of query terms while these last

works focus on the term importance in the query depending on the documents statistics and relevance or on terms' semantic similarities.

The related works in this paper, focused on two branches: (i) *Meaning making* of words in the query, by finding relations between words and conceptualizing words, using graphs and (ii) *importance of words* in the query, by adding weights. Some of these studies appear to have visible drawbacks while others have to be discussed further. In the following section, an explanation of the inconveniences of these last studies is proposed, where the inconveniences are related to the *query format* used in the IRS query formulation.

3. THE OUTCOME OF RELATED WORK

3.1 Concept Map Query

One of the differences between Concept Maps and Mind Maps is the use of linking phrases in the Concept Maps [19]. This means that the relation between every two nodes (concepts) in the Concept Map should be specified (i.e. labeled). However, in the Mind Map the relations are not specified by the user, they are abstract association in the mind of the user. In Information Retrieval, if we imagine the cognitive situation of a user who is specifying his information need, certain questions would evolve: (i) would this user be able to specify the kind of the connection between the words he is remembering? (ii) Can this user define his words of search as a relation (e.g. let the query: "The goal of development of diesel". "Goal" could be a relation between "development" and "Diesel") or as a term (or a concept)? Considering that the user is confused in the step of query formulation, it would be easier to mention only the words remembered and extracted from his mind naturally by association. Connections between the words in the brain of the user are interesting because they could define better the user thinking. In fact, as Tony Buzan expressed "without connections, thinking would not even exist" [36]. These connections should not be labeled (causality, composition, "is used", "is created by"...) since the user is formulating an idea not knowledge. Knowledge is acquired after learning, it expresses meaningful, solid and organized thoughts [19]. However an idea of search can be a messy or faulty thought. The Concept Maps are knowledge models, they could not be dedicated only for the purpose of query formulation.

3.2 Semantic Graph Query

Like a Concept Map, a semantic graph has concepts and relations between concepts, which would involve the same inconveniences as in the section 3.1. Also, the query formulated by the semantic graph in SyDOM does not mention any *level of importance* between concepts used in the query. It only mentions the kind of relations between concepts.

The outcome of the state of art suggests a query formulation where there are: (i) the *exploitation of the links* happening between remembered terms during user query formulation (ii) the non-labeling of links in the query and (iii) the need of discriminating terms according to their *importance for the user*. A user system is generally reticent to new ways asking to add supplementary information. For example, for IRS query, the user could be reticent to adding manually importance weights for the words in the query, etc. In order to reduce this reticence, a simple and faithful way of idea extraction from the user mind has to be suggested. We propose in the next section a modeling of queries based on Mind Mapping technique. This modeling technique is justified by its convenience with cognitive sciences' assertions about the information recall process from the human mind.

4. OUR APPROACH: MIND MAP QUERY

The information need is relative to a knowledge gap or to an incompleteness or inadequacy situation [37]. As a first step, an individual starts by searching in his internal memory (human memory), how to express this need. Then in a second step, he searches the missing knowledge in an external memory (documents). Considering the first step, cognitive sciences should be explored and used in a way that would accommodate the natural user's mind behavior in the process of query formulation.

4.1 Theoretical Foundations

The information representation in the brain is very complex. Many researches tried to simulate the knowledge representation and the human reasoning. We present in the following, the theories of thinking in the Brain, memory and mind that are agreed by the majority of cognitive scientists.

4.1.1 The Brain: a Detector of Models

Words and concepts are not saved in the human brain in an isolated way. The human brain models and constructs mental maps of information. Human brain organizes the raw data in schemes [17]. It detects models (i.e. structures that organize information) through certain habits of mind. It generates inductively new mental models of knowledge or concepts (i.e. a unit of knowledge) by interconnecting information [17]. According to Monroe and Pendergrass [38], the brain actively joins ideas together through the neurons networks, it deletes some information, joins others in associative schemes or it modifies the existing structures to give sense to new information in the brain. When the brain is confronted to a list of information, it is always trying to reconstruct unconsciously pieces and networks of information in a multitude of overlapped models [17]. This complex organism is partially directed by *sequential processes*. Cognitive sciences studies state that the mind contains linear and procedural knowledge, but the foundation of knowledge going from basic facts to decision making consists of non-linear models [13].

4.1.2 Human Memory: the Association Mechanism

According to Norman and Bobrow[38], the extraction of information from the brain is often triggered by matching a context with a saved concept in the human memory. The remembering of a concept will trigger the recall of other associated concepts. This process is called by Meyer and Schvaneveldt[38,39], the *spreading activation* mechanism (see an example in the Figure 2).

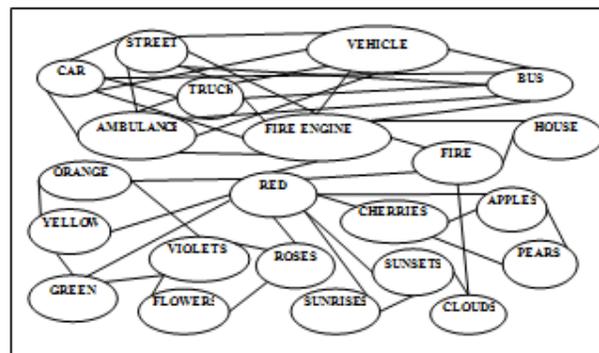


Figure 2. A stereotype of memory concepts representation [39]

The human memory can be divided in two kinds: The *episodic* and the *semantic memory*. According to Eric Jensen, the semantic memory is activated by mechanisms of association, similarity or contrast [17]. According to Capra [17], models existing in the human mind cannot be measured. They have to be extracted out of the mind, by mapping them. Cognitive scientists tried to extract from the mind several kinds of models: *Descriptive models, episodic models, causal models, sequential models, generalization/principal models, conceptual models*[40]. There are plenty of theories about these mental models in Cognitive Sciences defined in [40, 41] such as the *double-coding theory, connectionist models theory*, etc. We did not adopt one of these theories because they are argued by cognitive scientists. We consider in our theoretical background two phenomena: the detection of models by the brain, and the spreading activation of associations in the human memory. In the following section, we introduce our proposal of query formulation using these two cognitive findings.

4.2 Query Expression by User's Idea

An *Idea* is the result of the mind reflection process. It is a Greek word coming from the verb "*idein*" meaning "to see". An idea is then a tool for the perception of the mind [42]. A user's idea about a subject can change through time, since his mind can omit some details and remember others about the same subject at a specific time. This does not necessarily mean that the knowledge has been modified in the user's mind, but that the *mind recall process* depends on the user's situation and information need. Unlike *knowledge*, an idea can express the information differently according to the user, the situation and the information need.

The query formulation proposed in this paper, aims to make a link between the human brain and Information Retrieval, in order to offer to the user a deeper way for expressing his idea, and potentially reducing the *confusion* and the *omission of words* that the user faces in query formulation.

Our contribution is divided in two steps: (i) a Mind Map query formulation by the user and (ii) an internal representation of the Mind Map query by the IR system.

4.2.1 Query Formulation by Mind Maps

This step consists in representing the user query by a Mind Map. It allows the user to construct his own associations of ideas. The reasons for choosing Mind Maps are the following:

- *The association aspect*: The nature of links between ideas is not identified by users (no obligatory semantic hierarchy, no labeling of links). The links represent simply the associations between ideas that happen in the mind of the user while thinking.
- *The graph aspect*: It represents the radiant aspect of the brain, considering that the brain does not remember the ideas in a list [16].
- *The relative importance of terms*: In the thinking process, the user could accord level of importance to his thought: (a) words representing the central idea and (b) words coming by associations from the mind, which could be important or not very important. The words which are important are mentioned near to the central idea and the others are mentioned far from it.

We illustrate in the following examples, these three different aspects.

Illustration 1: In this example, we illustrate the necessity of the migration from the "bag of words" approach to the Mind Maps modeling approach. Let a user information need: "*The definition of a*

concept in a thesaurus or by standard norms". The bag of words query in IRS would be for example: "*definition of concept thesaurus standard norms*".

The problems are as follows:

- The system cannot guess the nodes in the brain of the user. Furthermore, for an IRS using techniques of *popular terms* reweighting, or *high weighted terms* in documents, the first documents retrieved would consider more these terms in the query. So, if we assume that "thesaurus" is popular or high weighted term in documents, then the first documents retrieved would deal more with the term "thesaurus" than "definition of concept". The *user importance degree* is absent in the query. The user information need is poorly expressed and controlled by one aspect of information need expression: the terms.
- Considering that in the classical IRS the terms in the query have all the same importance, the user could hesitate which term to add to the query among many others.

We illustrate a representation of the query by a Mind Map (see Figure 3), which tries to solve the problems mentioned above. The benefits of the use of Mind Map query are as follows:

- The associative aspect of terms allows the user to mention the words that comes to his mind with a new dimension of query control: The terms' weights of his idea. In fact, as a contrast with a linear textual list, a Mind Map represents the relative importance of different ideas (by computing their distance with the central idea and the heights between each other).

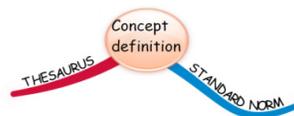


Figure 3. An example of Mind Map query

- The system understands that a user considers a term as not very important in the query. So, the user can mention the remembered terms without worrying about a misunderstood or a loss of his central idea of search through the query words.

Illustration 2: Let an information need of a query be "a good java course which describes inheritance and polymorphism, but also contains other notions of java. This course should contain exercises". If this query is formulated in a bag of words "good java course inheritance polymorphism exercises", we are as a user gambling with the IRS so it understands the terms (or concepts) importance in the query. In fact, the IRS can return java courses that describe only "polymorphism and inheritance" while the Information need is about a "good java course" not only "a java course about inheritance and polymorphism". The information need can be formulated by a Mind Map such as in the Figure 4.

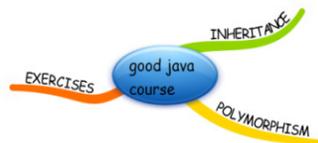


Figure4. A Mind Map example

Illustration 3: Let the idea of search "the definition of a semantic resource, for example thesaurus, ontology". The user in this query specifies that he is searching for documents that define "what is

a semantic resource?” and mentions some examples of semantic resources to get closer to documents that deepen the subject (see Figure 5). If this query is mentioned in a bag of words, the user would probably avoid mentioning the keywords “ontology” and “thesaurus” that comes to his mind as examples and would formulate instead the query “semantic resource definition”. The user would avoid expressing his whole idea in a bag of words form, in order to diminish IRS misunderstanding.

We present in the following section the internal interpretation of Mind Maps queries, by defining the formula used for nodes levels in the user’s idea of search.

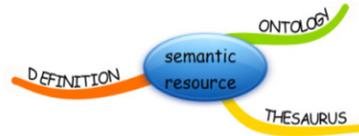


Figure5. A Mind Map example

4.2.2 The Internal Semantic for Mind Map Queries

In classical IRS, the number of occurrences of a term in a query is equal to 1. The weighting of query terms is uniform. These weights reflect the absence of term importance for the user. In our contribution, we consider the user terms’ importance by a graph with levels of importance: The more a remembered idea is far from the central idea of search, the less the term is important. In a quantitative way, terms in the first level of the Mind Map are considered σ times more important for the user than associated words in the second level. The weights of terms (w_i) in the graph are traduced in the IRS by this proposed formula :

$$w_i = \sigma^{(h-p_i-1)} \times a \quad (\text{Error! Bookmark not defined.})$$

Where :

σ : The power of importance between levels ($\sigma > 1$)

a : The weight attributed to the leafs of the graph (it could be Boolean or equal to TF or another weighting formula such as in [32])

p_i : The depth of the node n_i in the query graph

h : The height of the query graph

The proposed measure (3) could be classified in the relative importance semantic of query weighting formulas [28]. The calculus of this measure is illustrated in the following example.

Illustration 4: Let an information need: “Documents about Precision measure in Information Retrieval, for example: GMAP, MAP”. It could be expressed by a Mind Map query (see Figure 6).



Figure6. A Mind Map example

The four nodes (n_i) of the query are:

n_1 : Precision, n_2 : MAP, n_3 : GMAP, n_4 : Information Retrieval. Their respective weights are w_1, w_2, w_3, w_4 . They are calculated as follows, where we suppose that $\sigma = 2$ and $a=TF$ (Term frequency):

$$w_1 = 2^{(2-0-1)} \times TF = 2 \times \left(\frac{1}{5}\right) ; w_2 = w_3 = w_4 = 2^{(2-1-1)} \times TF = \frac{1}{5}$$

In this example, the weight of “Precision” term is twice important than the weights of “MAP”, “GMAP”, “Information Retrieval” terms.

To summarize, our approach offers these advantages:

- The *indirect* expression of term *importance* by the user
- The expression of the information need according to the *mind functioning* (spreading activation mechanism): For some cases, the user may not be able to define exactly his need but he may have information (related words) which could help in better expressing himself. So, the user can use associative terms (such as related examples or field of study) to get closer to his need. With the bag of words approach, the specification of associated terms to the central need is often misleading, because the IRS could not probably guess the central need and the related words.
- The expression of the user’s idea in a more structured format than the bag of words, which can facilitate the user thinking process while formulating his need.

To prove the accuracy of our approach, we present in the following, experimentation that we elaborated on a Medical collection test.

5. EVALUATION

We have conducted experiments in order to see the impact of Mind Map query on the retrieval accuracy. In order to fulfill this aim, we implemented into a classic retrieval component, the query weighting formula (see formula (3)) traducing the Mind Map in the Retrieval process.

5.1 The Experimental Environment

5.1.1 The Test Collection Features

We elaborated experimentation onto the test collection CLEF2009 (Cross Language Evaluation Forum) within the medical image retrieval track [45].The corpus collection test contains 74’902 images from 20’000 English journal articles in Radiology. There are 25 queries in the collection test. Both queries and documents are composed of images which are described by an XML text caption. Following our purpose, we used only the text form. The collection CLEF2009 also proposes for the queries, three languages: English, German and French. We opt for the English language.

5.1.2 The IRS Implementation Features

We explain in this part, the specific parameters in the IRS with which the tests were elaborated. The internal model of knowledge representation used in the IRS is the *vector space model*. In the IRS, we used a semantic indexing. The indexing process was elaborated via the *MetaMap* analyzer which uses the UMLS Meta-thesaurus for concepts extraction. After the concepts extraction step,

the formula CF-IDF [46] is used for weighting the documents (queries) concepts. The correspondence between a document and a query is calculated with the cosine correlation measure. In our work, there were 16'514 indexed documents and 25 indexed queries from CLEF2009 collection. In our approach, we intervene in the indexing process only by changing the weights of queries concepts. So, the IRS will have only one changed parameter: The weighting formula of mind map queries.

5.1.3 Methodology of Evaluation

We have considered, in our experiments, Mind Maps with two levels: The first level expresses the central idea of search, and the second one contains associated ideas (it is possible to consider more than two levels in the Mind Map query, if future experiments on users demonstrate the need for it). Moreover, we make the supposition that each node in the Mind Map corresponds to one concept from the indexed query.

Considering that the queries in the test collection are in a linear form (bag of words), the counting of all the possible cases of Mind Maps is necessary. The result of this counting is a set of possible Mind Map queries. The similarity between documents and these possible queries is elaborated. The Mind Map query returning the best results in our approach is then compared to the results of the IRS using a bag of words query (Linear query).

In this experimentation, we tested the importance degree offered within Mind Maps according to hypotheses on the weighting formula (see formula (3)). We describe the hypotheses as follows:

Hypothesis 1: The importance weights of concepts in the query are different from each other, but the whole query weight remains the same as in the bag of words query. We assume that the variable a of the formula (3) is determined by the equation (4).

$$\sum_{i=1}^n w_i = 1 \quad (\text{Error! Bookmark not defined.})$$

Where: w_i is the weight of a node i in the mind map query and n is the number of nodes in the query

Hypothesis 2: A node in the query is σ times more important than another node in the query. Values of σ (min: 2, max: 5) are tested. In order to not lose the significance of a node in the Mind Map query ($w_i \rightarrow 0$), we opt for the maximum $\sigma = 5$. In fact, the more the value of importance σ increases the less the weights of the nodes with a depth $p_i = 2$ are significant especially if the number of nodes are important (n).

The number of experiments to be performed depends on the factors in the Information Retrieval System (IRS). The table 1 illustrates the factors to experiment and their low and high levels.

Table 1. The extreme levelsofthe IRS factors

Factors / Two Levels	Low Level	High Level
Power of importance (σ)	2	5
Height of the query graph (h)	1	2
Default weight of a node (a)	$\frac{1}{n}$	$\frac{1}{n \sum_{i=1}^n \sigma^{(h-p_i-1)}}$

We outperformed 3 experimental runs combining some of the factors and levels (see table 2). The two first experimental runs correspond to our Mind Map IRS and the last one corresponds to the classic IRS.

Table 2. Experimental runs of the IRS

Power of importance (σ)	Height of the query graph (h)	Default weight of a node (a)
2	2	$\frac{1}{n \sum_{i=1}^n \sigma^{(h-p_i-1)}}$
5	2	$\frac{1}{n \sum_{i=1}^n \sigma^{(h-p_i-1)}}$
2	1	$\frac{1}{n}$

5.2 Experimental results

To evaluate our approach, we used three measures: (i) The MAP measure for comparing the global performance of IRSs, (ii) the Precision-Recall couple for evaluating the behavior of IRS and (iii) values of precisions at different levels of documents retrieved (P@n) for evaluating the quality of the first n retrieved results.

We present the results of an IRS based on Mind Map query and an IRS based on bag of words query. In the 25 queries of the collection test, 7 queries are mono-concept (a Mind Map with one node) and 18 queries are multi-concepts. We focus on the results of the 18 multi-nodes queries of the collection test (multi-nodes Mind Maps).

Table 3. The global impact of Mind Map query

Queries set	Importance value σ	MAP		Δ MAP
		Mind Map queries	Linear queries	
18	$\sigma = 2$	0.2452	0.1801	+36.14%
	$\sigma = 5$	0.2542		+41.14%
25	$\sigma = 2$	0.2378	0.1909	+24.56%
	$\sigma = 5$	0.2443		+27.97%

We observe in the table 3 an augmentation of the MAP by 36.14% when the central node of the Mind Map is twice more important than the associated nodes and an improvement of the MAP by 41.14% when the central node is five times more important than the associated nodes. This global improvement of precision encourages the use of the proposed weighting formula of the Mind Map queries in an IRS.

Table 4. The average impact of the Mind Map query on the first documents retrieved (P@5)

Queries set	Importance value σ	P@5		$\Delta P@5$
		Mind Map queries	Linear queries	
18	$\sigma = 2$	0.4888	0.4333	+12.80%
	$\sigma = 5$	0.4555		+5.12%
25	$\sigma = 2$	0.536	0.496	+8.06%
	$\sigma = 5$	0.512		+3.22%

In table 4 and table 5, for the 5 and 10 first documents retrieved, there is an amelioration of precision for Mind Map queries. Due to this observation, modeling queries by Mind Maps could be employed in *precision-oriented systems* such as Medical IRS. We can see that the value of importance $\sigma = 2$ showed better results for the MAP and for the precision at the 5 first retrieved documents of the IRS better than the value $\sigma = 5$. However the value $\sigma = 5$ showed better results for the precision at the 10 first retrieved results than the value $\sigma = 2$.

Table 5. The average impact of the Mind Map query on the first documents retrieved (P@10)

Queries set	Importance value σ	P@10		$\Delta P@10$
		Mind Map queries	Linear queries	
18	$\sigma = 2$	0.4388	0.3722	+17.89%
	$\sigma = 5$	0.4666		+25.36%
25	$\sigma = 2$	0.5	0.4520	+10.61%
	$\sigma = 5$	0.52		+15.04%

Figure 7 shows the positive impact of modeling queries by Mind Maps, on the performance of an IRS. In fact, the curve of Mind Map approach is always higher than the curve of the classical approach either for the value $\sigma = 2$ or the value $\sigma = 5$.

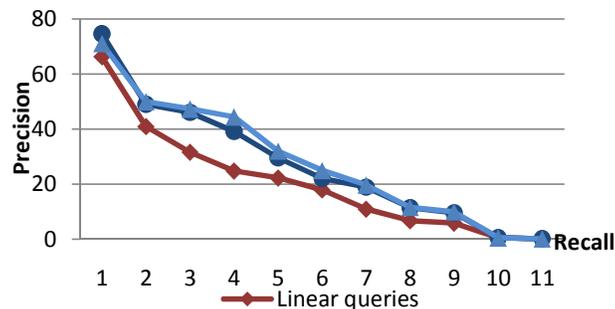


Figure 7. Precision/Recall curves (18 queries)

According to the results, the incorporated importance weights in the Mind Map queries proved to give better accuracy than the uniform weighting of bag of words queries (Linear queries). The

values of importance σ should be further tested and adjusted according to experiments on real users Mind Map queries in an IRS. Moreover, user studies appear to be fundamental in order to perceive the IRS users behavior during query formulation by Mind Maps.

6. CONCLUSION AND PERSPECTIVES

In IR query formulation process, the user tries to find the best way to express his need, by telling the remembered words. These words clarify the user's idea of search. This clarification could not be faithfully fulfilled if all words are specified in an unrelated way (bag of words). We suggested in this work to find a more faithful way for user's query formulation. We depicted from the Human brain's organizing: the *detection of models mechanism* and from the Human memory recall: the *associations' mechanism*. We suggested for IRS query formulation a cognitive map which satisfies these last mechanisms: a Mind Map. Also, we incorporated the importance degree in order to let the user express the important words and the less important ones in the Mind Map query.

Demonstrations on IR users should be elaborated in a further work, in order to study more the relation between associations (in human mind) and levels of importance in a query. Some critics to this work would be about guessing the association of the central idea of search by the system (not by the user). These critics are interesting but in our opinion, we need first to explore and understand better the expression of the users' search idea (by Mind Maps) before the deducing of the user's mind associations by the system.

REFERENCES

- [1] Göker, A. and Davies, J. (2009) Information Retrieval: Searching in the 21st Century. John Wiley & Sons Ltd.
- [2] Spink, A., Jansen, B., Wolfram, D. and Saracevic, T. (2002) From E-Sex to E-Commerce: Web search changes. IEEE Computer, 35, 3 (March. 2002), 107-109.
- [3] McGee, M. (2010) Google Weighs In on Query Length. Small Business Search Marketing. www.smallbusinesssem.com/google-query-length/3273/
- [4] Trellian. (2013) Keyword and search engines statistics. keyworddiscovery.com/keyword-stats.html?date=2013-01-01
- [5] Kamvar, M., Kellar, M., Patel, R. and Xu, Y. (2009) Computers and iPhones and Mobile Phones, oh my!: a logs based comparison of search users on different devices. In Proceedings of the 18th International Conference on World Wide Web (Madrid, Spain, April 20-24, 2009). WWW'09. ACM, New York, NY, 801-810.
- [6] Song, Y., Ma, H., Wang, H. and Wang, K. (2013) Exploring and Exploiting User Search Behavior on Mobile and Tablet Devices to Improve Search Relevance. In Proceedings of the 22nd International World Wide Web Conference (Rio de Janeiro, Brazil, May 13-17, 2013). WWW'13.ACM, New York, NY, 1201-1212.
- [7] Bhatia, M. and Akshi, K. (2010) Paradigm shifts: from pre web information systems to recent web-based contextual information retrieval. Webology, 7, 1 (Jun. 2010).
- [8] Sparrow, B., Liu, J. and Wegner, D.M. (2011) Google Effects on Memory: Cognitive Consequences of Having Information at Our Fingertips. Science 333, (Aug. 2011), 776-778.
- [9] Carr, N. (2010) What the internet is doing to our brains: The shallows. W.W.Norton & Company Inc., New York, NY
- [10] Hyerle, D. (2008) Thinking Maps : A visual language for learning. In Knowledge Cartography. Thinking Foundation.
- [11] Emberling, D. (2005) About Cognitive Maps. Developmental Consulting Inc. www.developmentalconsulting.com/pdfs/About_Cognitive_Maps_vA.pdf
- [12] Dagan, R. (2002) Cognitive Mapping. Intraspec. <http://intraspec.ca/cogmap.php>

- [13] Hyerle, D. and Piercy, T. D. (2010) Thinking Maps: The Cognitive Bridge to Literacy A Visual Language for Bridging Reading Text Structures to Writing Prompts. In Thinking maps : a new language for learning. Thinking Foundation, USA.
- [14] Hyerle, D. (2008) Thinking Maps: Visual Tools for Activating Habits of Mind. In Learning and Leading with Habits of Mind. ASCD.
- [15] Echaned Learning. Graphic Organizers. www.gmbservices.ca/Jr/GraphicOrganizers.htm.
- [16] Buzan, T. and Buzan, B. (1993) The Mind Map Book: How to Use Radiant Thinking to Maximize Your Brain's Untapped Potential. Plume, New York.
- [17] Hyerle, D. (2009) Visual Tools for Transforming Information into Knowledge. Corwin Press.
- [18] Topic Scape. Mindmaps Directory. www.topicscape.com/mindmaps
- [19] Novak, J. D. and Cañas, A. J. (2008) The theory underlying concept maps and how to construct and use them. Technical Report. Institute for Human and Machine Cognition (IHMC)
- [20] Boughanem, M. and Baziz, M. (2006) An IR model based on a sub-tree representation. In Leading the Web in Concurrent Engineering: Next Generation Concurrent Engineering, P. Ghodous, Ed. IOS Press, 450-457.
- [21] Leake, D., Maguitman, A., Reichherzer, T., Cañas, A. J., Carvalho, M., Arguedas, M. and Eskridge, T. (2004) "Googling" from a concept map : towards automatic concept-map-based query formation. In Proceedings of the 1st International Conference on Concept Mapping (Pamplona, Spain, September 14-17, 2004). 409-416.
- [22] Kwon, S. Y. (2006) The comparative effect of individually-generated vs. collaboratively-generated computer-based concept mapping on science concept learning. Doctoral Thesis. Texas A&M University.
- [23] Roussey, C., Calabretto, S. and Pinon, J.M. (2001) A Multilingual Information Retrieval System for Digital Libraries. In Proceedings 5th East European Conference on Advances in Databases and Information Systems (Vilnius, Lithuania, September 25-28, 2001). ADBIS'01. 98-111.
- [24] Mohidin, F. (2010) The Google Wonder Wheel and Mind Maps. Mind Map Tutor. www.mindmaptutor.com/2010/05/the-google-wonder-wheel-and-mind-maps/
- [25] PremiumSEOsolutions. (2012) Google KnowledgeGraph: New search technique. www.premiumseosolutions.com.au/blog/seo-news/google-knowledge-graph-new-search-technique/
- [26] Chuck, F. (2011) Pearltrees extends its mind mapping and curation application to the iPad. The Mind Mapping Software Blog. mindmappingsoftwareblog.com/pearltrees-for-ipad/
- [27] Bordogna, G. and Pasi, G. (1993) A Fuzzy Linguistic Approach Generalizing Boolean Information Retrieval : A Model and Its evaluation. J Am. Soc Inform. Sci. 44,2 (Sept. 1992), 70-82.
- [28] Zadrozny, S. and Kacprzyk, J. (2009) On an interpretation of keywords weights in information retrieval : some fuzzy logic based approaches. Int. J. Uncertain. Fuzz. 17, 1 (Aug. 2009), 41-58.
- [29] Herrera-Viedma, E., Lopez-Herrera, A. G., Alonso, S., Porcel, C. and Cabrerizo, F. J. (2007) A Linguistic Multi-level Weighted Query Language to Represent User Information Needs. In Proceedings of the IEEE International Conference on Fuzzy Systems (London, UK, July 23-26, 2007). FUZZY-IEEE'07. IEEE, 1-6.
- [30] Herrera-Viedma, E., Alonso, S., Cabrerizo, F. J., Lopez-Herrera, A. G. and Porcel, C. (2007) A software tool to teach the performance of Fuzzy IR systems based on weighted queries. In Proceedings of the 1st International Workshop on Teaching and Learning of Information Retrieval (London, UK, January 10, 2007). TLIR'07.
- [31] Arif, A.S.M., Rahman, M.M. and Mukta, S.Y. (2009) Information Retrieval by modified term weighting method using random walk model with query term position ranking. In Proceedings of the International Conference on Signal Processing Systems (Singapore, May 15-17, 2009).
- [32] Singhal, A. (1996) Term Weighting Revisited. Doctoral Thesis. Cornell University.
- [33] Monz, C. (2007) Model Tree Learning for Query Term Weighting in Question Answering. In Proceedings of the 29th European Conference on IR Research (Rome, Italy, April 2-5, 2007). ECIR'07. Springer. 589-596.
- [34] Wang, B., Zhou, Y., Zhang, Q. and Xuanjing, H. (2011) Learning the Weight of the Query Term from the Relevance Feedback. In the 7th International Conference on Natural Language Processing and Knowledge Engineering (Tokushima, Japan, November 27-29, 2011). NLP-KE'11. IEEE. 43-50
- [35] Zheng, W. and Fang, H. (2010) Query Aspect based Term Weighting Regularization in Information Retrieval. In Proceedings of the 32nd European Conference on IR Research (Milton Keynes, UK, March 28-31, 2010). ECIR'10. Springer. 344-356.

- [36] McAdam, T. (2010) Maximize the Power of Your Brain Using Mind Mapping. Self Improvement Information. www.selfimprovementinformation.com/tony-buzan-maximize-the-power-of-your-brain-using-mind-mapping/
- [37] Ingwersen, P. (1996) Cognitive Perspectives of Information Retrieval Interaction : Elements of a Cognitive IR Theory. *J. Doc.* 52, 1 (Mar. 1996), 3-50.
- [38] Hickie, K.M. (2006) An Examination of Student Performance in Reading/Language and Mathematics after Two Years of Thinking Maps Implementation in Three Tennessee Schools. Doctoral Thesis. East Tennessee State University.
- [39] Collins, A.M. and Loftus, E.F. (1975) A spreading-activation theory of semantic processing. *Psychol. Rev.* 82, 407-428.
- [40] Marzano, R.J., Pickering, D.J. and Pollock, J.E. (2001) Classroom instruction that works: Research-based strategies for increasing student achievement. McRel, USA
- [41] Schraw, G. Knowledge Representation. (2011) www.education.com/reference/article/knowledge-representation/
- [42] Rockmore, T. (2010) Kant and Phenomenology. The University of Chicago.
- [43] Ceusters, W., Smith, B. and Goldberg, L. (2005) A Terminological and Ontological Analysis of the NCI Thesaurus. *Method. Inform. Med.* 44, 4 (2005), 498-507.
- [44] IT Informaters. Common Mind Maps. www.informationtamers.com/WikIT/index.php?title=Common_mind_maps
- [45] ImageCLEF. ImageCLEF 2009 medical retrieval task. imageclef.org/2009/medical
- [46] Goossen, F., Ijntema, W., Frasinca, F., Hogenboom, F. and Kaymak, U. (2011) News Personalization using the CF-IDF Semantic Recommender. In Proceedings of the International Conference on Web Intelligence, Mining and Semantics (Sogndal, Norway, May 25-27, 2011). WIMS'11.
- [47] Ussery, B. (2008) Google–Average Number of Words Per Query have Increased!. Bleu Blog. www.beussery.com/blog/index.php/2008/02/google-average-number-of-words-per-query-have-increased/

AUTHORS

Rihab Ayed is a first year PhD student in Tunisia Polytechnic School, University of Carthage, Tunisia, actually specialized in Graph search issues in Information Retrieval Systems. She graduated in 2009 with a 'Computer Science Applied on Management' Bsc degree, and an 'Information Systems' Msc degree in 2011. She studied during her Msc thesis the most natural way to convert the human mind idea to a query understandable by Information Retrieval Systems.



Farah Harrathi is a post doctorate researcher specialized in Information Retrieval and more precisely in Semantic and Multilingual Indexing. He obtained his PhD degree in 2009 from the INSA Lyon, France. He has been an associate professor at Higher Institute of Multimedia Arts of Manouba, University of Manouba and visitor professor at the Faculty of Sciences of Tunis and at the Faculty of Sciences of Gafsa, Tunisia.

Mohamed Mohsen Gammoudi is currently a full Professor at Higher Institute of Multimedia Arts of Manouba, University of Manouba, Tunisia. He is responsible of SCO-ECRI team in Research Laboratory RIADI. He obtained his habilitation to supervise research in 2005 at the Faculty of Sciences of Tunis. He got his PhD in September 1993 in Sophia Antipolis Laboratory I3S/CNRS. Professor Gammoudi's professional work experience began in 1992 when he was assigned as an assistant at the Technical University of Nice. Then he was hired as a visiting professor between 1993 and 1997 at Federal University of Maranhao, Brazil. Since, he has supervised several PhD and master thesis.



Mahran Farhat is currently a PhD student at the Faculty of Sciences of Tunis (FST), he is a member of Laboratory RIADI. He obtained his master in 2011 at the Faculty of Sciences of Tunis (FST). He started his PhD in 2012 under the leadership of Professor Mohamed Mohsen Gammoudi. Currently, he served as a contractual assistant professor at ESEN, University of Manouba.

