Visualizing relationships in Hierarchical Small World Networks

Antoun Yaacoub[†] , Ali Awada[††] and May Dehayni[††],

IRIT - University of Paul Sabatier, Toulouse, France

Lebanese University, Faculty of Sciences, Hadath, Lebanon

Summary
This research provides a method for studying and visualizing different types of relationships in various worlds such as dictionaries and Web pages. As a first step, we identify different types of relationships between objects of the universes previously mentioned. The second step consists in developing a method to quantify the relationship to study. The constituents of the universe (Web pages, items associated with entries in a dictionary ...) form a HSWN graph (hierarchical small world network) whose nodes are entities and edges reflect a direct link (hypertext, definitional ...) between two nodes. This leads to introduce a method for studying the structure of large graphs of hierarchical small world type. Our approach is based on the use of Markovian matrices. A matrix is multiplied k times in order to quantify the relationship between all the nodes of the graph. To illustrate this approach, examples and results on Web and linguistic graphs are given.

Key words:
Graph, Hierarchical small world network, link structure, synonymy, visualization.

1. Introduction

The research in lexical semantics rely increasingly on large electronic resources (electronic dictionaries, ontologies …) from which one can obtain various semantic relations between lexical units. These relationships are naturally modeled by graphs. Although they describe very different lexical phenomena, these graphs have in common a unique feature: they are organized as a hierarchical small world network (HSWN). These networks are characterized by a small diameter, a distribution of degrees that follows a power law and a high local density clustering. We have studied the structure of these graphs for the lexicon and the Web in order to show the organization which is implicitly encoded in their structure.

The construction of semantic networks to model the connections between the different items of a given universe is an old problem. Ross Quillian in 1966 had laid the foundations of this field by proposing a semantic network for knowledge representation and reasoning [1]. Recent studies in this field have led to the development of tools to visualize and manipulate very large graphs with thousands of nodes. The display of such graphs is based on algorithms for placing items in order to respect some "aesthetic" construction rules.

2. Related works

In this section we describe some types of relationships in dictionaries and Web pages universes. We start by defining these relations then we summarize some works in this research field.

2.1 Relationships

The semantics [2] is a branch of linguistics that aims to study the meaning of words. Several works in computational linguistics have studied the relationships among dictionary entries. These works have considered the dictionary as a graph where words are represented by vertices and relationships between words by an arrow. Several kinds of relationships between two words exist [3] [4]. Important semantic relations are:

- Synonymy (A denotes the same as B): it can be defined as a ratio of semantic proximity between words. The semantic proximity indicates that words have meanings very similar in a given context and that two synonyms are necessarily similar
- Antonymy (A denotes the opposite of B)
- Hyponymy (A is subordinate of B; A is kind of B)
- Hyperonymy (A is superordinate of B)
- Meronymy (A is part of B)
- Holonymy (B has A as a part of itself)
- There are many other types of relationships such as eponymy, metonymy, synecdoche, catachresis, metaphor, pantonymy, paronymy ...

The other type of relationships that will take into consideration is the hyperlink in the World Wide Web (WWW). Analyzing the hyperlink structure between Web pages can provide an effective way to solve the problem of

Manuscript received June 5, 2010
Manuscript revised June 20, 2010

partitioning (clustering) of Web pages [5] [6]. This allows associating a page to some meanings (keywords) and may be helpful during the process of information search using a search engine.

Bibliometrics [7] is the study of written documents and their reference structure. Two similarity functions emerge from studies and bibliometrics are bibliographic coupling (due to Kessler [8]) and co-citation (due to Small [9]). For a pair of documents p and q, the first quantity is equal to the number of documents cited by p and q together and the second quantity is the number of documents that cite both p and q. Co-citation has been used as a measure of the similarity of Web pages by Larson [10] and by Pitkow and Pirolli [11]. Weiss et al. [12] define link-based similarity measures that generalize co-citation and bibliographic coupling to allow arbitrarily long chains of links.

Several methods have been proposed in this context to produce clusters from a set of nodes annotated with information similarity. Small and Griffith [13] use a breadth-first search to calculate the connected components of graph in which two nodes are connected by an edge if they have a positive value of co-citation. Pitkow and Pirolli [11] apply this algorithm to study the relationships based on the links in a collection of Web pages.

In the context of information retrieval, Deerwester et al. [14] propose the Latent Semantic Indexing method that involves applying a centroid scaling approach to a vector space model of documents [15]. This approach allowed to represent words and documents into a single common space of low dimension.

2.2 Lexical graphs "geometrization"

Ploux and Victorri [16] study cliques in lexical graphs. A clique is an undirected graph such that every two vertices are connected. A clique is a good candidate to form clusters in a graph. The high rate of clustering in a HSWN ensures the presence of a large number of cliques. Moreover, cliques have been used in the exploration of synonymy graphs in order to group each word with its synonyms [17].

Gaume et al. [18] [19] [20] propose a stochastic method for measuring and mapping the structures of local and global HSWN. The *Prox* method consists in transforming a graph into a Markov chain whose states are the graph vertices.

Muller et al. [21] present a method to exploit the structure of semantic graphs and calculate distance between words. They assume that the average distance covered by a particle between two nodes is an indication of the semantic distance between these nodes.

Awada and Chebaro [22] introduce the concept of "synonymetry" measurement as the proximity of meaning between two verbs of a language. This measure allows detecting and eliminating the metaphorical uses of verbs. They define the "N-connexity" as a new mathematical criterion to group synonymous verbs: a sub-graph forms an N-connexe component if each of its vertices is connected with at least N vertices of same sub-graph. They reformulate the notion of synonymy by considering that two words are synonyms if they belong to the same N-connexe component in dictionary graph.

Awada and Dehayni [23] propose a method to solve the polysemy problem by dividing the synonyms of a verb into groups called meaning components, each corresponding to a sense of the verb. Two verbs belong to a meaning component if there is, at least, a number of circuits whose length does not exceed a given threshold.

2.3 Visualization Interfaces

Visualization interfaces study the visual representation of large collections of information and use graphics to help understanding and analyzing data [24]. They deal with abstract data which do not have a geometric structure, such as unstructured text or points in a huge space [25] [26].

Various classifications of visualization interfaces have been proposed as in [27] which deals with information retrieval visualization or in [28] and [29] which perform a taxonomy of different visualization techniques. In the field of documents research and classification, Hearst [30] has developed a tool for presenting the keywords distribution of a query in the found documents. Cougar [31] allows the user to view all documents in relation to the searched themes. In Vibe [32], the user can view the documents in relation to terms of interest. Another interesting approach is that proposed by Cugini et al. through the interface Three-Axis Keywords Display [33]. This interface allows viewing words or combinations of words on orthogonal axes.

Other visualization techniques aim to highlight the relationships between different documents. The most common approach is the visualization of documents through a network as proposed by the search engine *Kartoo* in which documents are related according to the terms they have in common [34]. The classification of documents can group similar documents thereby reducing the number of items to display. Indeed, documents are no longer represented independently but as classes of documents [35] [36].

In the Web domain, different approaches have been proposed including the visualization tool *H3* [37] which is based on a hyperbolic viewer. *Walrus* is a tool for interactively visualizing large directed graphs in three-dimensional space. It is best suited to visualizing moderately sized (a few hundred thousand vertices) graphs

that are nearly trees. *Walrus* uses 3D hyperbolic geometry to display graphs under a fisheye-like distortion. By bringing different parts of a graph to the magnified central region, the user can examine every part of the graph in detail. Walrus was developed by Young Hyun at CAIDA based on research by Munzner [38]. *Vogue* is a Huge Hierarchies visualization tool based on the fractal approach [39]. The self-similarity that characterizes fractal allows users to visually interact with a huge tree in the same way at every level of the tree.

3. Functional architecture

In this section we propose a solution for each of the dictionary and Web pages universes. We explain how to build a HSWN corresponding to a semantic network and how to visualize relationships. We present an illustrative example and two case studies, one on the world of dictionaries and the other on Web pages.

Our solution is based on the use of a transition matrix (M) raised to the power k (M^k). The element $M^k(i, j)$ is the probability to reach the point j starting from the point i by a path of length k. Thus, the resulting matrix with coordinates in R^p (p is the number of nodes-entries in the matrix) contains information calculated on the whole graph and could be represented in R^2 after a Main Component Analysis (MCA) that keeps only the two first axes.

3.1 Proposed solution for dictionaries

In our survey, we use the WordNet ontology developed by linguists of the Cognitive Science Laboratory at Princeton University [40]. Its purpose is to identify, classify and link in various ways the semantic and lexical content of the English language. The core component of the system is the synset (synonym set), a group of interchangeable words, denoting a sense or a particular purpose. Each synset denotes a different sense of the word in question.

Fig.1 shows the functional diagram corresponding to the dictionary universe module. It uses the Tree-Tagger annotator to pick up the dictionary words. Then, it involves WordNet in order to build the corresponding HSWN graph after choosing the parsing relation (synonymy, antonymy …). The obtained graph is transformed into a N^p matrix and then into a Markovian one (in R^p) using Matlab. A set of k multiplication (k is introduced by the user) is applied to the obtained Markovian matrix in order to group words according to their relations in the graph. Then, we apply a MCA allowing to pass from R^p to R^2 in order to visualize the relation as a 2D cloud of points.

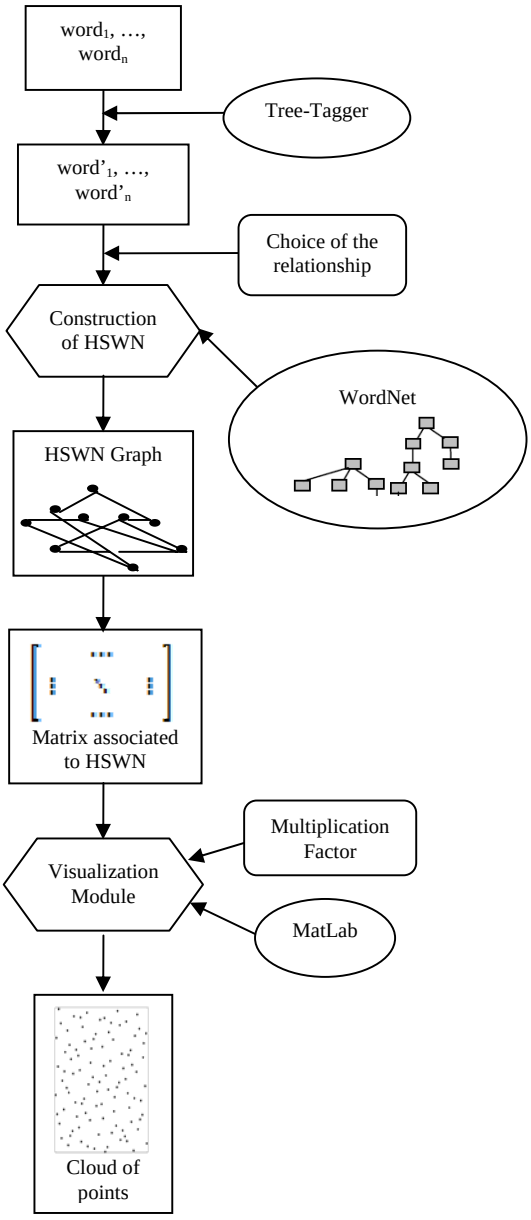

Fig. 1 The dictionaries universe: Functional diagram.

3.2 Proposed solution for Web pages

This module uses a crawler to retrieve all the pages linked to a given Web site in order to construct the corresponding HSWN graph. Thus, the same above described process for the dictionaries is applied to visualize the links as a 2D cloud of points (see Fig. 2).

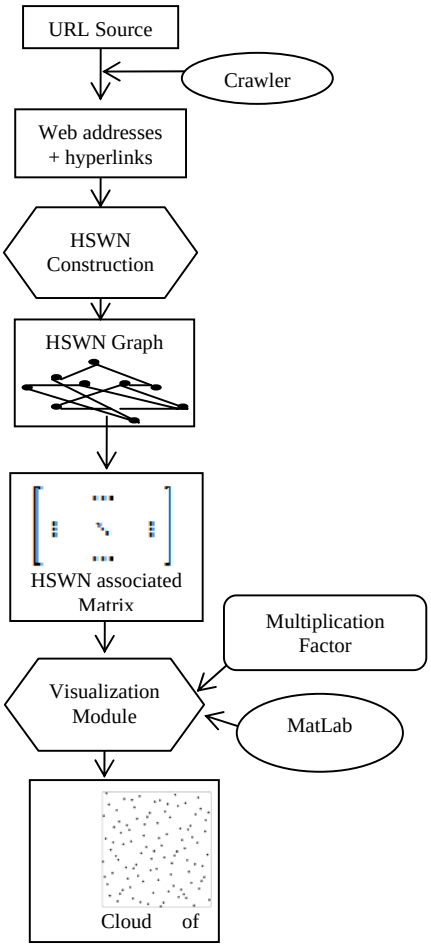

Fig. 2 The Web pages universe: Functional diagram.

4. Construction of HSWN

In this section we depict how to represent a semantic network corresponding to a dictionary and a set of Web pages as a HSWN graph. The resulting structure is transformed using a MCA. This method allows studying multidimensional data and is relevant to visualize the graphs.

4.1 Dictionary universe

Let a set of words $S=\{m_1, m_2, \ldots, m_n\}$ and a specific grammatical function (noun, verb, adjective or adverb). We pick a word m_i $(1 \le i \le n)$ from S, in order to find all its synonyms of the same grammatical function using WordNet.

For each new found synonym, we reiterate the same process until reaching the set of all words $\{m_2,\ldots,m_n\}$. This is possible because each graph of different grammatical functions in WordNet constitute a single connected component.

Afterwards, we build a square symmetric matrix M in which $M[i,j]$ depicts a synonymy relationship between the words m_i and m_j.

We should mention that we process a breadth-first traversal of the WordNet synonymy graph in order to prevent a quickly divergent path risk due to the existence of polysemic words in a path (see Fig. 3).

Example: If the user wants to study the synonymy relationship between the words: "dark", "shadow" and "night". The below figure depicts the corresponding WordNet semantic network.

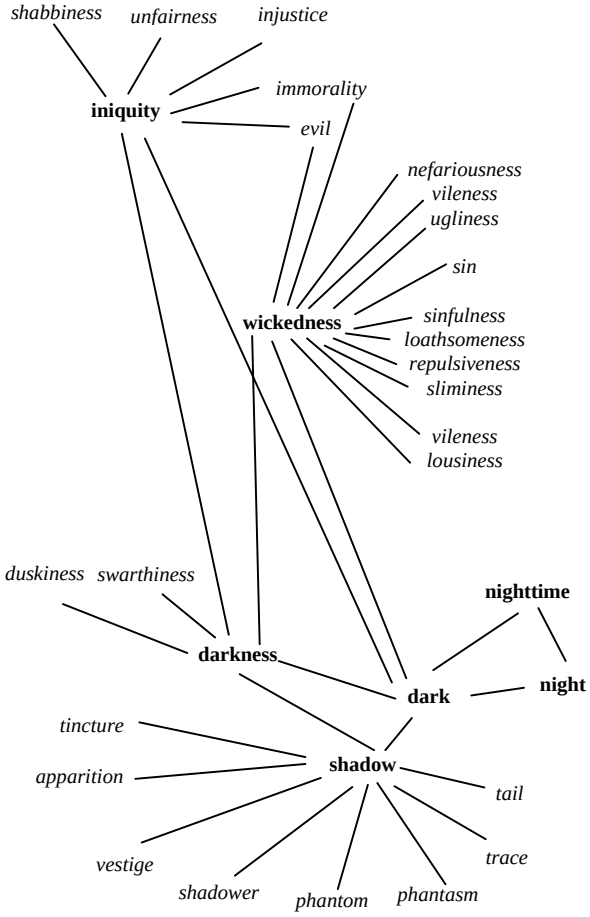

Fig. 3 "Dark" semantic network. Synonyms are in bold.

Since the relationship is the synonymy, there is more chance to get good synonyms during a breadth-first graph traversal because the synonyms of a word are in his close neighborhood. Thereby, as we move away from the initial word, the probability of finding synonyms declines.

4.2 Web pages universe

The same approach is used to build HSWN corresponding to the Web pages universe. Indeed, once the crawler searches all hyperlinks of a given pivot Web page, every referenced page is analyzed in the same above way for dictionary universe.

5. Visualization of relationships

Multi-factorial analysis methods provide the best possible information summary of data contained in a large table or matrix. Depending on the phenomena to be studied and the nature of the available data, an adequate multi-factorial method is applied. Indeed, there are several multi-factorial analysis methods, all based on the same mathematical theories [41]. Among these, we mention MCA, correspondence factorial analysis, multiple correspondence analysis …

In our approach, we adopt MCA that allows studying multidimensional numerical data and showing links between them. Data are represented as a cloud of points in a geometric space. The goal is then to find some sub-spaces that best represent the initial cloud of points.

6. Experiment and results

6.1 Illustration on a small graph

Since our application handles a huge number of data, a realistic example would be hard to understand by the reader. Thus, we choose to illustrate our approach on a small graph for a better comprehension.

Consider a graph G of 13 nodes and a set of edges reflecting a relation among them. The graph is depicted as follow:

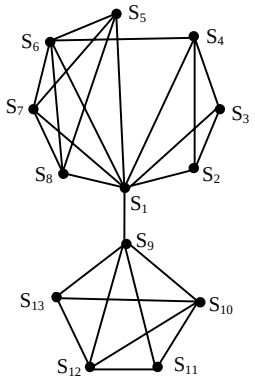

Fig. 4 Example of a small graph G

It is obvious that the component S_1 cannot be grouped with S_9 because there is no possible cycle between them. Here is the corresponding adjacency matrix A.

$$A = \begin{bmatrix} 0 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 0 & 0 & 0 & 0 \\ 1 & 0 & 1 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 1 & 1 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 1 & 1 & 1 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 & 0 & 1 & 1 & 1 & 0 & 0 & 0 & 0 & 0 \\ 1 & 0 & 0 & 1 & 1 & 0 & 1 & 1 & 0 & 0 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 & 1 & 1 & 0 & 1 & 0 & 0 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 & 1 & 1 & 1 & 0 & 0 & 0 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 1 & 1 & 1 & 1 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 1 & 1 & 1 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 1 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 1 & 1 & 0 & 1 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 1 & 0 & 1 & 0 \end{bmatrix}$$

Fig. 5 Adjacency matrix of G

In the corresponding Markovian matrix M depicted below, we notice that the sum of every line is equal to 1 because its values denote probabilities.

$$M = \begin{bmatrix} 0 & 1/8 & 1/8 & 1/8 & 1/8 & 1/8 & 1/8 & 1/8 & 1/8 & 0 & 0 & 0 & 0 \\ 1/3 & 0 & 1/3 & 1/3 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 1/3 & 1/3 & 0 & 1/3 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 1/4 & 1/4 & 1/4 & 0 & 0 & 1/4 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 1/4 & 0 & 0 & 0 & 0 & 1/4 & 1/4 & 1/4 & 0 & 0 & 0 & 0 & 0 \\ 1/5 & 0 & 0 & 1/5 & 1/5 & 0 & 1/5 & 1/5 & 0 & 0 & 0 & 0 & 0 \\ 1/4 & 0 & 0 & 0 & 1/4 & 1/4 & 0 & 1/4 & 0 & 0 & 0 & 0 & 0 \\ 1/4 & 0 & 0 & 0 & 1/4 & 1/4 & 1/4 & 0 & 0 & 0 & 0 & 0 & 0 \\ 1/5 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1/5 & 1/5 & 1/5 & 1/5 & 1/5 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1/4 & 0 & 1/4 & 1/4 & 1/4 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1/3 & 1/3 & 0 & 1/3 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1/4 & 1/4 & 1/4 & 0 & 1/4 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1/3 & 1/3 & 0 & 1/3 & 0 \end{bmatrix}$$

Fig. 6 Markovian matrix of G

Let us compute M^7 since in 7 edges we can reach all S_1 component nodes at least once, and slightly more than once for S_9 component nodes.

S_1	S_2	S_3	S_4	S_5	S_6	S_7	S_8	S_9	S_{10}	S_{11}	S_{12}	S_{13}
0.178	*0.071*	*0.071*	*0.095*	*0.094*	*0.118*	*0.094*	*0.094*	0.056	0.035	0.026	0.035	0.026
0.189	*0.080*	*0.081*	*0.107*	*0.095*	*0.122*	*0.095*	*0.095*	0.045	0.024	0.018	0.024	0.018
0.189	*0.081*	*0.080*	*0.107*	*0.095*	*0.122*	*0.095*	*0.095*	0.045	0.024	0.018	0.024	0.018
0.190	*0.080*	*0.080*	*0.104*	*0.097*	*0.125*	*0.097*	*0.097*	0.043	0.022	0.017	0.022	0.017
0.189	*0.071*	*0.071*	*0.097*	*0.106*	*0.131*	*0.106*	*0.106*	0.042	0.021	0.016	0.021	0.016
0.189	*0.073*	*0.073*	*0.100*	*0.105*	*0.129*	*0.105*	*0.105*	0.041	0.021	0.016	0.021	0.016
0.189	*0.071*	*0.071*	*0.097*	*0.106*	*0.131*	*0.106*	*0.106*	0.042	0.021	0.016	0.021	0.016
0.189	*0.071*	*0.071*	*0.097*	*0.106*	*0.131*	*0.106*	*0.106*	0.042	0.021	0.016	0.021	0.016
0.090	0.027	0.027	0.035	0.033	0.041	0.033	0.033	**0.161**	**0.146**	**0.111**	**0.146**	**0.111**
0.071	0.018	0.018	0.022	0.021	0.026	0.021	0.021	**0.183**	**0.169**	**0.127**	**0.169**	**0.127**
0.071	0.018	0.018	0.023	0.022	0.027	0.022	0.022	**0.185**	**0.170**	**0.123**	**0.170**	**0.123**
0.071	0.018	0.018	0.022	0.021	0.026	0.021	0.021	**0.183**	**0.169**	**0.127**	**0.169**	**0.127**
0.071	0.018	0.018	0.023	0.022	0.027	0.022	0.022	**0.185**	**0.170**	**0.123**	**0.170**	**0.123**

Fig. 7 Matrix M^7

Note that the presence of two components: S_1 component (in italic) and S_9 component (in bold).

0.148	0.055	0.055	0.074	0.074	0.092	0.074	0.074	0.092	0.074	0.055	0.074	0.055
0.148	0.055	0.055	0.074	0.074	0.092	0.074	0.074	0.092	0.073	0.055	0.073	0.055
0.148	0.055	0.055	0.074	0.074	0.092	0.074	0.074	0.092	0.073	0.055	0.073	0.055
0.148	0.055	0.055	0.074	0.074	0.092	0.074	0.074	0.092	0.073	0.055	0.073	0.055
0.148	0.055	0.055	0.074	0.074	0.092	0.074	0.074	0.092	0.073	0.055	0.073	0.055
0.148	0.055	0.055	0.074	0.074	0.092	0.074	0.074	0.092	0.073	0.055	0.073	0.055
0.148	0.055	0.055	0.074	0.074	0.092	0.074	0.074	0.092	0.073	0.055	0.073	0.055
0.148	0.055	0.055	0.074	0.074	0.092	0.074	0.074	0.092	0.073	0.055	0.073	0.055
0.148	0.055	0.055	0.074	0.074	0.092	0.074	0.074	0.092	0.074	0.055	0.074	0.055
0.148	0.055	0.055	0.073	0.073	0.092	0.073	0.073	0.092	0.074	0.055	0.074	0.055
0.148	0.055	0.055	0.073	0.073	0.092	0.073	0.073	0.092	0.074	0.055	0.074	0.055
0.148	0.055	0.055	0.073	0.073	0.092	0.073	0.073	0.092	0.074	0.055	0.074	0.055
0.148	0.055	0.055	0.073	0.073	0.092	0.073	0.073	0.092	0.074	0.055	0.074	0.055
0.148	0.055	0.055	0.073	0.073	0.0924	0.073	0.073	0.092	0.073	0.055	0.073	0.055
A	B	B	C	C	D	C	C	D	C	B	C	B

Fig. 8 Matrix M^{100}

Fig. 8 depicts the following partition: A={S_1}, B={S_2, S_3, S_{11}, S_{13}} C={S_4, S_5, S_7, S_8, S_{10}, S_{12}} and D={S_6, S_9}
Starting from any node and enough traversing the graph, we have a probability of about 0.148 to reach S_1. For B elements the probability is about 5.5% (4 nodes x 5.55% = 22.2%), 7.4% for C elements (6 x 7.4% = 44.4%), and 9.25% for D elements (2 x 9.25% = 18.5%). In Individual percentage, the ranking in descending order is A, D, C, B. This can be interpreted by the fact that S_1 is an important articulation node (a hub), then come more modestly S_6 and S_9. These classes of values characterize the "hub" property of a node but definitely not its membership to the same component in its class.
The next step is the application of a MCA in order to visualize these results in 2D.

X	Y	Node
-10.314	-0.0028072	S1
3.0949	-0.010477	S2
3.0949	-0.010477	S3
0.4132	-0.010041	S4
0.41322	-0.010341	S5
-2.2685	-0.0098616	S6
0.41322	-0.010341	S7
0.41322	-0.010341	S8
-2.2704	0.019808	S9
0.41145	0.017579	S10
3.0936	0.0098586	S11
0.41145	0.017579	S12
3.0936	0.0098586	S13

Fig. 9 MCA resulting matrix

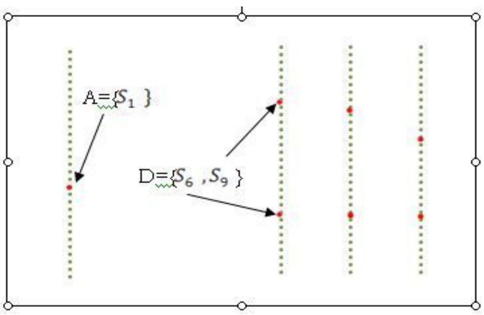

Fig. 10 Result visualization

Referring to the x axis coordinate in Fig. 10, we identify the same groups as Fig. 8, but we have more correlation on the y axis. So in terms of MCA, we keep the first two dominant factor axis in order to preserve all of the dispersion of the whole cloud of points.

6.2 Example on the dictionary universe

In this example, we study the synonymy relationship between “big” and “fat”.
The corresponding graph is composed of 264 nodes. Fig. 11 shows a spatial representation of these nodes and depicts a visual approximation of their semantic distance. The initial two words are marked by a circle on the graph. We have deliberately decided not to display words on the issues because several points overlap, leading to overload the figure and make it unreadable.

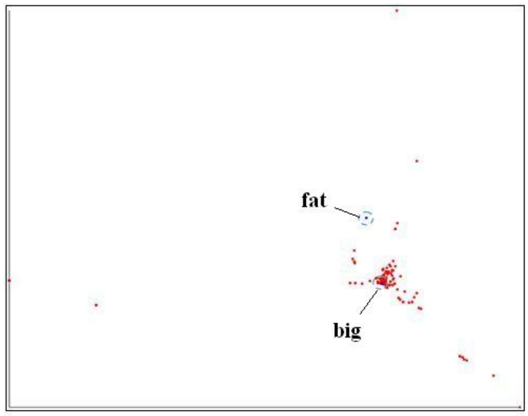

Fig. 11 Cloud of points corresponding to the matrix M

Fig. 12 provides more detailed information. In fact, after traversing 100 times the graph, the figure shows a potential path between nodes (generally the smallest one). Furthermore, the cloud of points that verify the synonymy relation is seen as a continuum between different words, allowing to switch from a meaning to another. Note that the figure confirms that the word “fat” is polysemic because it has a dense neighborhood. Indeed, we depict three groups of almost equidistant synonyms-nodes, each corresponding to a sense of the word fat.

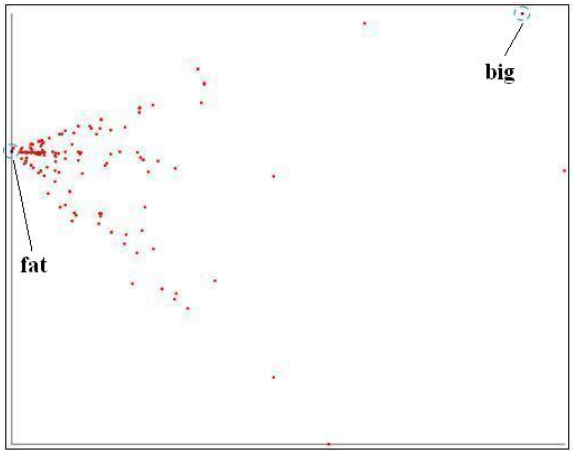

Fig. 12 Cloud of points corresponding to M^{100}

6.3 Example on the Web universe

This example corresponds to the http://antoun.yaacoub.org website study.
The graph contains 1637 nodes or Web pages. Fig. 13 shows a spatial representation of these nodes where the initial one is marked by a circle. We notice that the points in this cloud follow two orthogonal directions, expressing the lack of connection between the points of each direction with those of the other one.

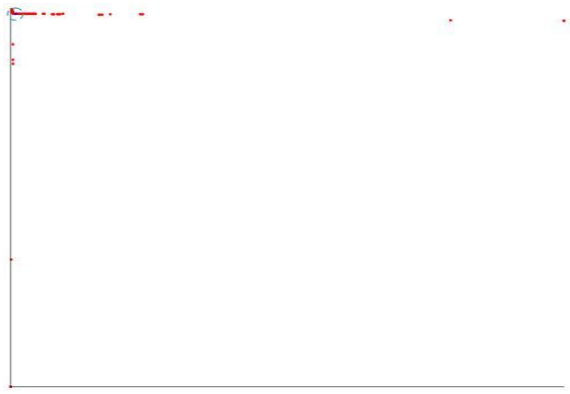

Fig. 13 Cloud of points corresponding to the matrix M

Fig. 14 provides more detailed information about the relationship between pages in each direction. In fact, after traversing the graph 100 times, the figure shows that the set of points arranged in the horizontal direction do not maintain a linear relationship (moving from one page to another requires more than one link). Finally, we notice a dense cluster (marked by an oval) corresponding to a highly interconnected set of pages.

Fig. 14 Cloud of points corresponding to M^{100}

Conclusion

In this paper, we have developed a graphical user interface (GUI) to explore relationships in dictionaries and Web pages universes. However, the induced graphs of these universes have in common unique features: they are organized as a HSWN. This property is verified when the studied relationship is the synonymy for dictionaries and the hyperlink for Web pages.

We have proposed an approach that allows visualizing a HSWN. The solution we proposed is based on the use of a transition matrix. Traversing the induced graph ‘x’ times is equivalent to raising the transition matrix to the power ‘x’. The resulting matrix contains information calculated on the whole graph (the multiplication of the matrix by itself gives information about the relationship among its nodes). This relationship is represented in R^2 after applying a MCA by keeping only the first two axes.

In our contribution, we validate our approach by comparing the obtained results with the theoretical ones. In the dictionary universe, the spatial representation of the nodes provides information on the synonymy relationship and a visual approximation of the semantic distance between words. Furthermore, our interface allows locating hubs and dense clusters on the graph and deciding whether a word is polysemic. In addition, a possible interpretation of multiple directions in the graph suggests that the number of branches corresponds to the number of meanings of the initial word.

Besides the ability of our GUI to explore different relationships in the universes mentioned above, it can generate files containing the computed results (the input nodes, the transition matrix, the MCA calculated coordinates ...). On the other hand, it supports files whose structure is adequate with a predefined. Our tool is

"generic" because it allows studying any universe that satisfies these conditions.

However, we were faced with purely technical issues. Indeed, the machines we have used are not adapted to sufficiently powerful calculations that require a large capacity of memory. This prevented us to test our approach on other universes and to propose more elaborated examples because the graphs size quickly explodes.

The possible perspectives of our work are about three areas:

- The first concerns finding an efficient way to build the matrix corresponding to the graph of synonyms from WordNet. Indeed, for a given word, its relationship with other synonyms in WordNet is a static Boolean (two words are synonyms or are not). So each word has a row in the matrix with lot of 0's and few of 1's. Since this relationship is fixed and does not change from one application to another, we may save place and time by storing these static matrices in a database or in sparse matrices.

- The second concerns coupling this tool to a search engine. Our tool can detect very dense clusters (highly connected pages or synonymous words). We could use this tool to provide a method for query expansion (using the dictionary universe component), and to combine the results obtained in the Web universe with *Hubs and Authorities* concepts to propose a more efficient information retrieval system.

- The third suggests a possible exploitation of our tool by developing a parameterized crawler on different criteria (link/tag type, traversal types ...) applied on structured documents (html, xml dictionary, text ...).

References

[1] M.R. Quillian. Semantic memory in semantic information processing, M.I.T. Press, 1968.

[2] J. Lyons. Eléments de sémantique, Larousse, 1978.

[3] G. A. Miller, R. Beckwith, C. Fellbaum, D. Gross, and K. Miller. Introduction to WordNet: An On-line Lexical Database, International Journal of Lexicography 3 (4), pp. 235 – 244, 1990.

[4] A. Lehmann, F. Martin-Berthet. Introduction à la lexicologie. Sémantique et morphologie, Lettres sup 2008.

[5] J.M. Kleinberg. Authoritative Sources in a Hyperlinked Environment, IBM Research. Report RJ 10076, May 1997.

[6] S. Chakrabarti, B. Dom, D. Gibson, J. Kleinberg, R. Kumar, P. Raghavan, S. Rajagopalan, A. Tomkins. Mining the Link Structure of the World Wide Web, IEEE Computer, August 1999.

[7] L. Egghe, R. Rousseau. Introduction to Informetrics, Elsevier, 1990.

[8] M.M. Kessler. Bibliographic coupling between scientific papers, American Documentation, 14, pp. 10-25, 1963.

[9] H. Small. Co-citation in the scientific literature: A new measure of the relationship between two documents, J. American Soc. Info. Sci., 24, pp. 265-269, 1973.

[10] R. Larson. Bibliometrics of the World Wide Web: An exploratory analysis of the intellectual structure of cyberspace, Ann. Meeting of the American Soc. Info. Sci., 1996.

[11] J. Pitkow, P. Pirolli. Life, death, and lawfulness on the electronic frontier, Proceedings of ACM SIGCHI Conference on Human Factors in Computing, 1997.

[12] R. Weiss, B. Velez, M. Sheldon, C. Nemprempre, P. Szilagyi, D.K. Giord. HyPursuit: A Hierarchical Network Search Engine that Exploits Content-Link Hypertext Clustering, Proceedings of the Seventh ACM Conference on Hypertext, 1996.

[13] H. Small, B.C. Griffith. The structure of the scientific literatures I. Identifying and graphing specialties, Science Studies 4, pp. 17-40, 1974.

[14] S. Deerwester, S. Dumais, T. Landauer, G. Furnas, R. Harshman. Indexing by latent semantic analysis, J. American Soc. Info. Sci., 41, pp. 391-407, 1990.

[15] C.J. van Rijsbergen. Information Retrieval, Butterworths, 1979.

[16] S. Ploux, B. Victorri. Construction d'espaces sémantiques à l'aide de dictionnaires de synonymes, Traitement automatique des langues, 39/1, pp.161-182, 1998.

[17] F. Venant. Polysémie et calcul du sens, Actes des 7es Journées internationales d'Analyse statistique des Données Textuelles (JADT), 2004.

[18] B. Gaume. Cartographier la forme du sens dans les petits mondes Lexicaux, JADT 2006, p 541-465. 2006.

[19] B. Gaume, L. Ferré. Représentation de graphes par ACP granulaire, actes d'EGC 2004 : 4èmes journées d'Extraction et de Gestion des Connaissances, Clermont Ferrand, 20-23 Janvier 2004.

[20] B. Gaume, K. Duvignau, O. Gasquet. Forms of Meaning, Meaning of Forms, Journal of Exprimental and Theoretical Artificial Intelligence, 14(1), pp. 61-74, 2002.

[21] P. Muller, N. Hathout, B. Gaume. Synonym Extraction Using a Semantic Distance on a Dictionary. Workshop on TextGraphs, at HLT-NAACL 2006, pp. 65–72, 2006.

[22] A. Awada, B. Chebaro. Etude de la synonymie par l'extraction de composantes N-connexes dans les graphes de dictionnaires. JEL2004, Nantes, France, 2004.

[23] A.Awada, M.Dehayni. Grouping dictionary Synonyms in sense components, Journal of Theoretical and Applied Information Technology, vol.6, No.1, July 2009.

[24] S.G. Eick. Graphically displaying text, Journal of Computational and Graphical Statistics, vol 3, pp. 127–142. 1994.

[25] T. Munzner. Guest Editor's Introduction IEEE Computer Graphics and Applications, Special Issue on Information Visualization, Jan/Feb 2002.

[26] S.K. Card, J.D. Mackinlay, and B. Shneiderman, Readings in Information Visualization: Using Vision to Think, Morgan Kaufmann Publishers, 1999.

[27] O. Zamir. Visualization of search results in document retrieval systems, General Examination, University of Washington, 1998.

[28] .H. Chi. A taxonomy of visualization techniques using the data state reference model, INFOVIS'2000, Salt Lake City, pp 69-75, October, 2000.
[29] M. Hascoët, M. Beaudouin-Lafon. Visualisation interactive d'information, Revue Information-Interaction-Intelligence (I3), « A Journal in Information Engineering Sciences », 1(1), 2001.
[30] M.A. Hearst. TileBars: visualization of term distribution information in full text information access, ACM Conference on Human Factors in Computing Systems (SIGCHI), Denver CO, pp. 59-66, May, 1995.
[31] M.A. Hearst. Using categories to provide context for full-text retrieval results, Conférence Recherche d'Informations Assistée par Ordinateur (RIAO): Intelligent Multimedia Information Retrieval Systems and Management, Rockefeller University, NY, October, 1994.
[32] K.A. Olsen, R.R. Korfhage, K.M. Sochats, M.B. Spring, J.G. Williams. Visualization of a document collection: the VIBE system, Information Processing and Management Journal (IPM), 29(1), pp. 69-81, 1993.
[33] J.V. Cugini, C. Piatko, S Laskowski. Interactive 3D visualization for document retrieval, Actes CIKM, Rockville MD, 1996.
[34] M. Chalmers, P. Chitson. Bead: explorations in information visualization, 15th International ACM SIGIR Conference on Research and Development in Information Retrieval, Copenhagen, pp. 330-337, 1992.
[35] O. Zamir, O. Etzioni. Grouper: a dynamic clustering interface to web search results, Computer Networks, vol. 31(11-16), pp. 1361-1374, May, 1999.
[36] T. Kohonen. Self-organised formation of topologically correct feature maps, Biological Cybernetics, vol. 43, pp. 59-69, 1982.
[37] T. Munzner. H3: Laying Out Large Directed Graphs in 3D Hyperbolic Space, Proceedings of the 1997 IEEE Symposium on Information Visualization, October 20-21 1997, Phoenix, AZ, pp. 2-10, 1997.
[38] T. Münzner. Exploring Large Graphs in 3D Hyperbolic space, 1998. http://www.cs.kent.edu/~jmaletic/cs63903/papers/Munzner98.pdf
[39] H. Koike. The Role of Another Spatial Dimension in Software Visualization, ACM Transaction on Information Systems, Vol. 11, No. 3, July 1993.
[40] G.A. Miller, R. Beckwith, C. Fellbaum, D. Gross, and K. Miller. Introduction to WordNet: An On-line Lexical Database, International Journal of Lexicography 3 (4), pp. 235 – 244, 1990.
[41] J.-C. Gower. Measures of similarity, dissimilarity and distance, Kotz S., Johnson N.-L. & Read C.-B. (eds), Encyclopedia of Statistical Sciences, vol. 5. New York: Wiley, pp. 397-405, 1985.

Antoun Yaacoub is currently a Ph.D. student in computer science in Université Paul Sabatier at Toulouse – France. He's conducting his research at the Institut de recherche informatique de Toulouse (IRIT) – France. His research focuses on defining, identifying and analyzing the flow of information (from a security point of view) in logic programs. He previously worked on various aspects of the French language and focused on the types of links existing between the words.

Ali Awada is an Associate Professor in computer science in the Faculty of Sciences of the Lebanese University, Hadath, Lebanon.
He previously worked on human/machine communication in natural language. He obtained a Ph.D. thesis from the Institut National Polytechnique at Toulouse – France. He is currently working in the domains of Artificial Intelligence and human/machine interface; more specifically ATMS, semantic distance, combining Attribute Grammars with human/machine communication, and using dictionary information in the process of request expansion in IRS.

May Dehayni is an Assistant Professor in computer science in the Faculty of Sciences of the Lebanese University, Hadath, Lebanon.
Her previous research and publications have been in the field of meta-modeling (MOF). Her Ph.D. thesis, in Université Paul Sabatier at Toulouse - France, proposes an approach of model transformation based on attribute grammars. She is currently working with Ali Awada in the domains of Artificial Intelligence and human/machine interface.