Identification and Extraction of Memes represented as Semantic Networks from Free Text Online Forums

Hector Beck-Fernandez[1,2], David F. Nettleton[1,3]

[1]Dept. Information Technology and Communications,
Universitat Pompeu Fabra, Barcelona, Spain
{hector.beck, david.nettleton}@upf.edu
[2]Área de Ingeniería en Computación e Informática,
Universidad de Tarapacá, Arica, Chile
[3]IIIA-CSIC, Bellaterra, Spain

Abstract. Memes have recently come into vogue in the context of 'viral' transmission of basic information units in online social networks. However, from their original general definition in a sociological context, there is still much work to be done from an information technology viewpoint. This includes such issues as how to process memes from real text corpus, formal definitions for knowledge representation, meme refinement and selection. In order to address these issues, in the current paper we adapt definitions from the semantic network and information retrieval fields to extract semantic network memes from free text documents, and then give some examples in the context of a simple online forum.

Keywords: Memes, semantic networks, information retrieval, free format text.

1 Introduction

As defined in a sociological context by Dawkins[1] and Blackmore[2], a meme is understood as a basic element of useful knowledge, or meta-information, which can transmitted from one individual to another. However, from an information technology point of view, many technical and implementation challenges remain, such as how to identify and extract key memes from free text document corpuses. The study of memes has a high potential utility for understanding and modelling information diffusion/influence in Online Social Networks and applications such as recommender systems [3]. Hence the work is motivated by the technical challenges on the one hand, and the application potential of the results, on the other.

In this paper the main focus will be on the problem of defining and identifying semantic network type structures in free text, which can then be used to represent memes. We use information retrieval and semantic networks concepts to identify and extract the key memes from a larger candidate set. A simple example is given of an online forum of comment posts to illustrate how the framework could be applied in practise.

The structure of the paper is as follows: in Section 2 we present the state of the art and related work; in Section 3 we present the definitions for the documents, semantic network concepts (entities and relations) and memes; in Section 4 we given some examples for the definitions of Section 3; in Section 5 we consider meme metrics and how we can use them to identify the 'top' memes extracted using the definitions and examples of Sections 3 and 4; finally in Section 6 we give the conclusions.

2 State of the Art and Related Work

The term "meme" was originally defined by Dawkins in [1], and has been recently applied to the study of how information spreads through Internet and OSNs.

According to Dawkins [1] a "meme" or a "memetype" is similar to a "gene" or "virus". It consists of a basic unit of information circulating among a community, and research from a social sciences perspective has studied how it serve as a mechanism to propitiate cultural and social evolution[4]. In [4], Heylighen and Chielens compare the 'meme' with the 'gene' and formalize the following memes properties: 'longevity', the duration that an individual meme survives; 'fecundity', the reproductive activity of a meme; 'copy-fidelity', the degree to which a meme is accurately reproduced.

In [5], Bordogna and Pasi propose a schematic definition for memes using an OWL schema, followed by the definition of several operators to extract memes from online blog posts using information retrieval methods and n-grams (contiguous sequences of n items from a given sequence of text). A fuzzy-type matching is performed to evaluate the fidelity of a given blog post to an original meme description. Finally, the longevity is considered by ordering the text entries by their timestamp and taking into consideration the fidelity.

Leskovec et al. [6] developed a framework for tracking short textual memes in an online news media environment, identifying a broad class of memes that exhibit a wide spread and rich variation on a daily basis. Simmons et al. [7] presented a study about meme mutation in social networks. They uncovered patterns in the rate of appearance of new variants, their length and popularity, and developed a simple model that is able to represents these attributes. Nettleton in [8] presents a wide-ranging survey of OSN analysis, covering themes such as ‘influence and recommendation’ and ‘information diffusion’, which includes contextual entity tracking using memes. Baydin and López de Mántaras [9] present an evolutionary algorithm based on the concept of memes. They used semantic networks to represent the individual pieces of information, and employed the 'genetic' concepts of crossover and mutation to model changes over time.

Now we will briefly summarize some of the literature with respect to the extraction of semantic networks from text. Szumlanski in [10] extracted semantic networks based on frequency and concept affinity, from Wikipedia texts using the WordNet [11] ontology database to identify related concepts. In [12], Jiang and Conrath's highly referenced paper describes a semantic similarity metric based on corpus statistics and a lexical taxonomy. They present an approach for measuring semantic similarity/distance between words and concepts which uses a distributional analysis of the corpus data. In [13], Chen et al. deal with the elicitation of semantic networks based on concepts relevant to the data mining of specific datasets. In [14], Kok and Domin-

gos present an unsupervised approach to extracting semantic networks from large volumes of text. They use the TextRunner system [15] to extract tuples from text, and then induce general concepts and relations from them by jointly clustering the objects and relational strings in the tuples. Their approach is defined in Markov logic using four basic rules to extract meaningful semantic networks.

3 Extraction of Semantic Network Memes from Free Format Text

In this Section we present the definitions for the meme environment (documents, concepts and relations) which will allow us to identify the key memes, represented as semantic networks, in a free format document corpus.

3.1 Introduction

There are two main processes: **(i)** which acts on the complete document set *D* to identify key concepts and relations. This is comprised of Definitions 1 to 3; **(ii)** which acts on individual documents to compact the semantic networks (eliminate redundant relations and identify the minimal semantic networks). This is comprised of Definitions 4 to 6.

Process 1: The objective of *Definition 1* is to identify the most relevant subset of documents and key concepts from the complete document corpus. Then, *Definitions 2* and *3* identify the relations between the key concepts. We note that *Definitions 1* to *3* act on the complete document corpus *D*.

Process 2: *Definitions 4*, *5* and *6* deal with eliminating redundant relations and finding the minimum semantic networks between concepts. We note that *Definitions 4*, *5* and *6* act on individual documents *d*.

We note the importance of the use of *thresholds* in the processing. The thresholds are determined statistically from the probability distributions of the corresponding metrics. The threshold can be defined by a quartile limit or by point of inflexion, from the corresponding distribution.

3.2 Definitions which Define the Extraction of Semantic Network Memes

Definition 1. A ***concept*** is an *n*-gram[1] (excluding *stopwords*[2], in the information retrieval sense) that is present in a significant number of documents in a document collection. Formally, let *D* be the total document collection. Then, an *n*-gram x_i is a ***concept*** when it satisfies the condition:

$$p_D(x_i) = \frac{\text{Nº of documents in } D \text{ which contain } x_i}{\text{Nº of documents in } D} = \frac{|D(x_i)|}{|D|} > \alpha \quad (1)$$

[1] Word or group of consecutive words where *n* is the number of words making up the sequence.

[2] Examples of lists of *stopwords* can be found at http://www.ranks.nl/resources/stopwords.html

where $\alpha \in [0,1]$ is a value known as threshold which is user defined. The threshold α indicates the percentage of documents containing an *n*-gram to be considered a concept. How is this value chosen? Low values of α will obtain many concepts; on the other hand, higher values of α will obtain fewer concepts. As we consider a document collection which is a free text comments forum, we are interested in those concepts that have most presence in the discussion. As an initial approximation, we could choose a moderately high value for α, in the order of 0.70 ± 0.05. Empirically, we could consider the three highest deciles in a frequency distribution table of candidates for concepts. Other definitions can be found in [16, 17].

In a given free text block written by a user of an online community, some concepts will be related to each other, in a way that has meaning for that community. Concepts such as "democracy", "is" and "participation" have little meaning when each is taken in isolation, however if they are related by means of a verbal expression (which may be a another concept), then they acquire much more meaning. For example, "participation is democracy". In this context we must determine which concepts are co-occurrences and which are related.

We recall that two concepts are a *co-occurrence* if they are at a distance less than n words, in the same sentence, or in the same paragraph. In the first case, a limit of 4 can be placed on the value of *n*; an interesting study on the co-occurrence of words is Ferrer-i-Cancho and Solé's paper[18]. The following Definition 2 provides a way to determine related concepts.

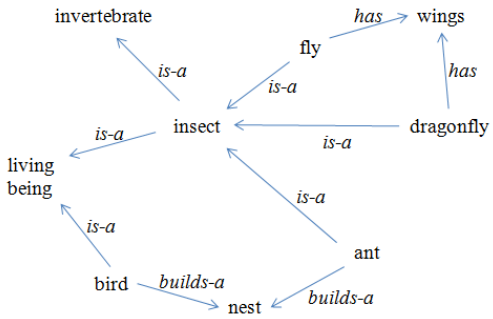

Fig. 1. Semantic network with 9 concepts and 3 distinct relationships (*has*, *is-a*, *builds-a*)

Definition 2. ***Relationship related*** (R): Let x_i and x_j be concepts in a document collection D, $x_i R x_j \Leftrightarrow p_D(x_i|x_j) > \gamma \wedge p_D(x_j|x_i) > \gamma$; where:

$p_D(x_i|x_j) = \frac{p_D(x_j,x_i)}{p_D(x_j)}$, and

$$p_D(x_j, x_i) = \frac{\text{Nº of documents in } D\text{; where } x_j \text{ and } x_i \text{ are co-occurrences}}{\text{Nº of documents in } D} = \frac{|D(x_i,x_j)|}{|D|} \quad (2)$$

That is, this is so if both concepts appear together in many documents in the collection D. In this case, the threshold $\gamma \in [0,1]$ indicates a measure to determine when both concepts are considered related, and is assigned by the user and verified empirically. Again, high values of γ could provide few concepts and low values of γ could

provide many related concepts. By extracting the concepts that correspond to verbs, nouns and adjectives, a syntactic and semantic representation of text can be represented in the form of semantic networks.

A semantic network (SN) is a notation that allows us to represent ideas with meaning and which represent knowledge. An SN is represented by a graph in which the nodes are concepts (nouns and adjectives) and the arcs are the relationships (verbal expressions) between them [19]. Fig. 1 shows a semantic network with 9 concepts and 3 relationships. This form of notation is used, for example, in the fields of natural language processing and information retrieval [20], among others.

Many kinds of relationships can be derived from a semantic network; we will only consider those which are relevant to our present work. As a starting point, we will use some of the definitions given in [17] by Oh et al., changing the notation and adapting them to the present context.

Definition 3. ***Relationship superset-subset***. If the document set of x_i $\left(D(x_i)\right)$ is almost included in the document set of x_j $\left(D\left(x_j\right)\right)$ then we say that $D\left(x_j\right)$ is a *superset* of $D(x_i)$ and we denote $D\left(x_j\right) \rightarrow D(x_i)$, formally:

$$D\left(x_j\right) \rightarrow D(x_i) \Leftrightarrow p_D\left(x_j|x_i\right) > \delta \text{ and } p_D\left(x_i|x_j\right) < \delta \quad (3)$$

In this case, $\delta \in (0,1)$ can be calculated empirically using the equation

$$|D(x_i)| * \delta < \left|D(x_i, x_j)\right| < \left|D(x_j)\right| * \delta \quad (4)$$

where $|D(x)|$ is the cardinality of set $D(x)$.

From definitions 1, 2 and 3 we have now identified the concepts as the most important terms in a document set. That is $x_i \rightarrow c_i$, when the given thresholds α, γ and δ are satisfied. In the following definitions 4 to 6 we will consider a given document d_q belonging to document set D.

Definition 4. ***Relationship redundant***. A relationship $d_q(x_i) \rightarrow d_q\left(x_j\right)$ is *redundant* if there exist one o more concepts such that $d_q(x_i) \rightarrow d_q(x_k) \rightarrow \cdots \rightarrow d_q\left(x_j\right)$ in a semantic network.

Definition 5. ***Closest Superset***. Let $\mathcal{C}_x = \left\{d_q(c_1), d_q(c_2), \ldots, d_q(c_k)\right\}$ the set of *Supersets* of x, i.e. $\mathcal{C}_x$ is the set of all $d_q(x_i)$ such that $d_q(x) \rightarrow d_q(c_1)$, $d_q(x) \rightarrow d_q(c_2)$, ... , $d_q(x) \rightarrow d_q(c_k)$. The *Closest Superset* of x is the smallest of all $d_q(c_i)$.

Definition 6. ***Minimal Semantic Network***. Let graph $\mathcal{G} = (\mathcal{C}, \mathcal{R})$ be a semantic network where $\mathcal{C}$ is a set of concepts and $\mathcal{R}$ is the set of relationships between the concepts. A semantic network $\mathcal{G}' = (\mathcal{C}, \mathcal{R}')$ with $\mathcal{R}' \subset \mathcal{R}$, is a *minimal semantic network* if, for all relationships $r_k = \left(c_i, c_j, Type\right) \in \mathcal{R}'$, $d_q(c_i)$ is the *closest superset* of $d_q(c_j)$

With respect to definition 6, we note that each relationship in a semantic network can be expressed by a triple $\left(x_i, x_j, type\right)$ where x_i and x_i are concepts and "*type*" is the type of relationship between x_i and x_j. In [17], Oh et al. proved that in a minimal semantic network the relationships between concepts are not redundant.

4 Examples

In this Section, with reference to Figs. 2, 3 and 4, we will give an example of each of the aspects we have described in Section 3. We note that the objective of the future work will be to automate the process as much as possible, however we envisage as semi-automatic scheme which may require some manual annotation of the original text and semi-supervised processing in other steps, such as in [18]. Although these implementation details are out of the scope of the current paper, we can say that in order to construct the semantic networks, we would need to distinguish between the entities and the relations from the initial set of concepts. This could be done using natural language processing software tools and a relationship-instance repository together with WordNet[11], *http://wordnet.princeton.edu/*, in order to identify entities (e.g. nouns, adjectives) and relations (e.g. verbs, adverbs).

Firstly, in Fig. 2 we see a simplified example of a typical online comments forum for a newspaper article. That is, a newspaper publishes an article about a given theme and below the article the registered users are allowed to post their opinions. What typically happens is that users with differing opinions create a debate in which some users state their opinions and other users either support or reject all or part of those opinions.

We observe in Fig. 2, that user 1 has posted a comment, which is replied by user 2. Then user 3 posts a new comment, which is replied by user 1, whose comment is in turn replied by user 2. We can clearly see that the central concepts are about foxes and dogs

Concepts: correspond to the search terms, which can be entities and/or relations. In Fig. 3, the semantic networks formed include the entity concepts 'fox', 'dog', 'brown', 'quick', 'lazy', and the relation concepts 'is', 'jumps-over', 'hunts'. As mentioned, semi-automatic tools exist for identifying syntax structures, however we must not underestimate the difficulty of correctly identifying the relations between entities, especially when a concept has different meanings dependent on the content. For example, the concept 'quick' can be a noun, adjective or adverb. For the present work, we assume a manual revision of the ambiguous cases. In Table 1 we see the concepts, their syntactic classification and the corresponding assignment as entity or relation.

Concept	**Possible syntactic categories**	**Chosen syntactic category**	**Entity or relation**
fox	noun, verb	noun	Entity
dog	noun, verb	noun	Entity
brown	noun, verb, adjective	noun/adjective	Entity
quick	noun, adjective, adverb	adjective	Entity
lazy	adjective	adjective	Entity
hunt	verb, noun	verb	Relation
jump	noun, verb	verb	Relation
over	noun, adjective, adverb	adverb	Relation
Is	noun, verb	verb	Relation

Table 1**.** Concepts, syntactic categories and assignments as entity or relation.

Documents: a document is a block of text (comment) written by a user. In information retrieval, if we formulate a query to search for a set of terms (or concepts), such as {fox, dog}, the query will return a set of documents in which one or more (depending if the query is AND or OR) of the query terms appears. Hence, a document will contain one or more concepts which are susceptible to be formed into one or more semantic networks. Hence in Figs. 2 and 3 we see there are five documents, designated as d_1 to d_5.

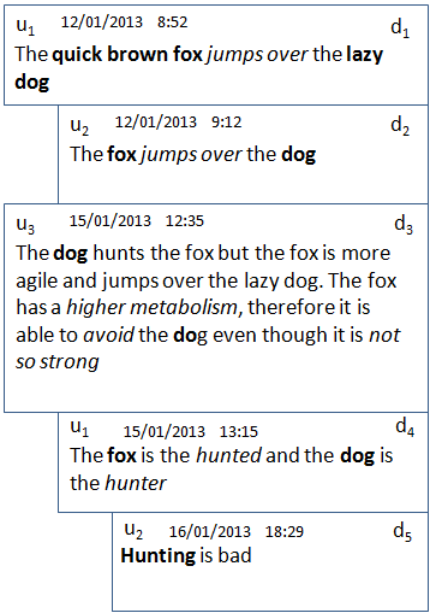

Fig. 2. Online forum example: user's posts, with date and timestamps

Semantic network (candidate meme): a semantic network is made up of two or more entity concepts which are related by one or more relation concepts. A document may contain one or more semantic networks, made up of corresponding concepts. In Fig. 3 we see that we have extracted three significant memes from all the potential semantic networks which can be constructed from the respective texts. Later, in Section 5 we will consider how we can use the meme metrics to identify the most significant memes.

Superset-subset: If a set of documents Sd_1 is included within another set of documents Sd_2 then Sd_1 is a subset of Sd_2 and Sd_2 is a superset of Sd_1. This is related to the information retrieval concept of document retrieval sets corresponding to queries made of one or more query terms. In the current context the query terms would be the concepts making up the memes, that is, each meme is a potential query. With reference to Fig. 3, consider the following example: the query {fox, dog, jumps} retrieves the set of documents $Sd_1 = \{d_1, d_2, d_3\}$; the query {fox, dog} retrieves the set of documents $Sd_2 = \{d_1, d_2, d_3, d_4\}$, which is a superset of document set Sd_1. Likewise, Sd_1 is a subset of Sd_2.

Redundancy: a relation (link) between two concepts is redundant if it already exists via another path. With reference to Fig. 4a, we see that the link between 'fox' and 'wolf' is redundant because it is already implicit (inherited) through the links between 'fox', 'dog' and 'wolf'.

Document set id	Query terms (entity concepts)	Document set returned by query
Sd_1	{fox, dog, jumps}	$\{d_1, d_2, d_3\}$
Sd_2	{fox, dog}	$\{d_1, d_2, d_3, d_4\}$
Sd_3	{fox, lazy, dog}	$\{d_1, d_3\}$

Table 2**.** Queries and document sets

Closest superset: the smallest superset with respect to a given subset. Returning to the example of Fig. 3, consider three queries, those we defined previously, Sd_1 and Sd_2, and a new one Sd_3 = {fox, lazy, dog} which returns documents $\{d_1, d_3\}$. Hence, the smallest superset with respect to Sd_3 will be Sd_1, as opposed to Sd_2, given that Sd_2 contains four documents whereas Sd_1 contains only three. In Table 2 we see the queries and the corresponding document sets.

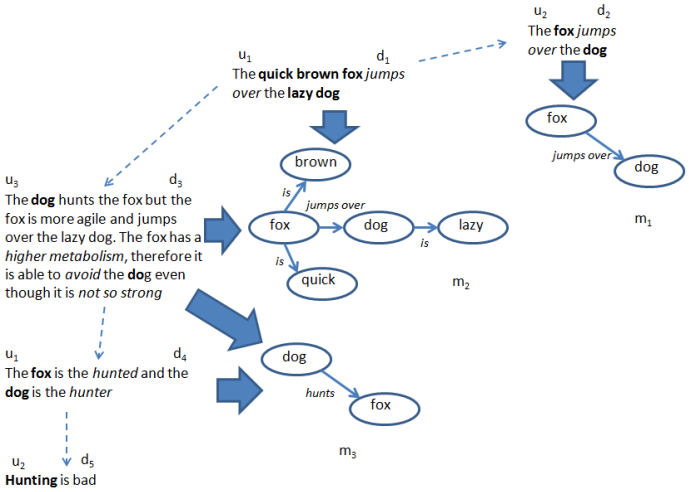

Fig. 3**.** Online forum example: documents (comment texts), users and memes

Compact: within a document, all groups of concepts (memes) are connected together by common concepts. With reference to Fig. 4b, we see that one unique semantic network has been formed by a (weak) link between two memes (concept groups with strong links).

Minimal: each group of concepts (meme) is separated from any other group of concepts. All links (relation concepts) are designated as being strong. With reference to Fig. 4b, we see that two distinct memes (concept groups) are identified.

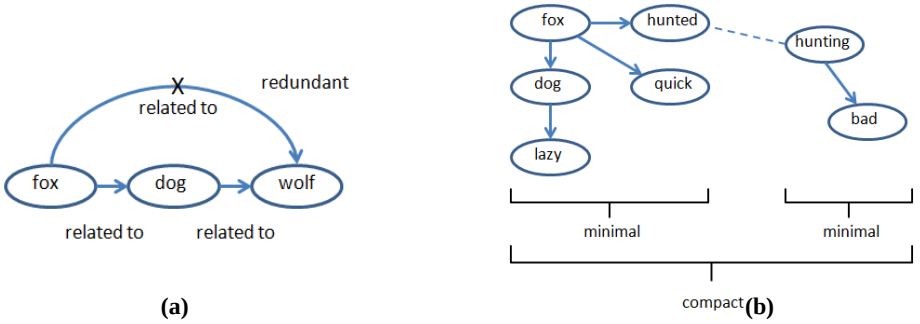

Fig. 4. (a) Example of a *redundant* relation in a semantic network; (b) Example of a *compact* and a *minimal* semantic network

Meme: is a semantic network which is composed of entity concepts with strong links (relation concepts), equivalent to the definition of 'minimal' (above). However, we apply further processing to identify the most relevant memes in a document collection, using the metrics of Section 5.

5 Incorporation of the Meme Metrics

In this Section we will first describe how we can use the three meme metrics of *longevity*, *fecundity* and *copy-fidelity*, to select the 'top' memes. Then we will give an example of processing using the memes depicted in Figs. 2 and 3. We note that we perform the meme metric based selection once the minimal memes have been obtained through the semantic network extraction process described in Section 3.

5.1 Meme Metric Based Selection Process

In order to perform the meme metric based selection, we represent the users as a directed graph, through which the memes are considered to 'move'. The implementation details of the graph and associated data structures are out of the scope of the present paper.

The selection process is performed in four steps: ***(i)*** obtain a value for each of the three metrics for each meme; ***(ii)*** obtain the distribution of the values of each metric for all memes; ***(iii)*** establish a cut-off point (threshold) for each metric based on their distributions; ***(iv)*** identify the memes which are above the thresholds for all metrics.

Step 1: obtain a value for each of the three metrics for each meme.

> ***1.1.*** Longevity L for a given meme m is designated as m_L. m_L is equal to the number of different arcs that are traversed in a period of time t. *Implementation:* this is a simple numerical calculation derived from the initial and maximum timestamps.

1.2. Fecundity *F* for a given meme *m* is designated as m_F. m_F is equal to the number of different vertices that are visited in a given period of time *t*. *Implementation:* this is a simple numerical calculation derived from the directed graph of the users.

1.3. Copy-fidelity *I* for a given meme *m* is designated as m_I. m_I is equal to the degree of 'loss of fidelity' of a meme over a given time period *t* or for a given number of arcs traversed, *at. Implementation:* a similarity comparison function, with an appropriate distance metric, will be applied to evaluate the fidelity of a given meme (at time *t*) with respect to the original meme (at time *0*).

Note 1, graph structure: the representation of the users and the meme transit between the users will require the implementation of the appropriate data structures and data processing procedures.

Step 2: obtain the ordered distribution of the values of each metric for all memes.

2.1. The distribution of the longevity values m_L for all memes will be a vector d_L. The distribution of the fecundity values m_F for all memes will be a vector d_F. The distribution of the copy-fidelity values m_I for all memes will be a vector d_I.

Step 3, establish a cut-off point (threshold) for each metric based on their distributions:

3.1. The threshold for the longevity distribution d_L will be designated as λ. The threshold for the fecundity distribution d_F will be designated as φ. The threshold for the copy-fidelity distribution d_I will be designated as σ.

Note 2, thresholds: there are different statistical techniques we can use to assign the thresholds λ, φ and σ based on the numerical distribution. For example, we can identify an inflexion point, or we can use the top *x*% percentile, or use a supervised optimization technique. This process could be manual, automatic or semi-automatic.

Step 4, identify the memes which are above the thresholds for all metrics:

4.1. **Meme characteristics.** Consider a meme *m* whose characteristics *mc* are embodied as: concept entities $\{e_1, ..., e_n\}$, concept relations $\{ r_1, ..., r_m\}$, longevity m_L, fecundity m_F and copy-fidelity m_I.

4.2. **Meme threshold based selection.** *MT*(*mc*, λ, , φ , σ) is a meme threshold selection function, whose inputs for a given meme are the meme's characteristics, mc, as defined in *Step 4.1*, and the three thresholds as obtained from *Steps 1.1* to *3.1*. The output of function *MT* will be a binary value [0,1] for which 1 signifies that meme *m* is within all three thresholds and 0 signifies that it is not. We note that we could relax the meme threshold restrictions, to require only two, or just one threshold to be complied with.

5.2 Example of Meme Metric Based Selection

In this Section we will give a simple example of the meme threshold based selection, with reference to the memes m_1, m_2 and m_3 shown in Figs. 2 and 3. We note that, in

this example, time is measured as the number of arcs traversed, and not the difference between the timestamps.

Applying *Step 1* we obtain:

- Meme longevity: $m_{L1} = 3$, $m_{L2} = 1$, $m_{L3} = 2$
- Meme fecundity: $m_{F1} = 2$, $m_{F2} = 0$, $m_{F3} = 1$
- Copy-fidelity: $m_{I1} = 3$, $m_{I2} = 1$, $m_{I3} = 1$

Applying *Step 2* we obtain the distributions for each metric:

$d_L = \{3, 2, 1\}$; $d_F = \{2, 1, 0\}$; $d_I = \{3, 1, 1\}$

Applying *Step 3 we* establish the threshold for each metric distribution:

$\lambda = 3$; $\varphi = 2$; $\sigma = 3$

Applying *Step 4.1* we assign the meme characteristics for memes m_1, m_2 and m_3, respectively:

mc_1({fox, dog}, { jumps-over}, 3, 2, 3) ; mc_2({fox, dog, brown, quick, lazy}, {is, jumps-over}, 1, 0, 1) ; mc_3({fox, dog}, { hunts}, 2, 1, 1).

Finally, applying *Step 4.2* identifies the meme(s) which are above the thresholds for all metrics:

$MT(mc_1, \lambda, , \varphi , \sigma) = 1$; $MT(mc_2, \lambda, , \varphi , \sigma) = 0$; $MT(mc_3, \lambda, , \varphi , \sigma) = 0$.

Hence, m_1 is the only meme which is above all three thresholds and is therefore selected as the top meme based on the metric thresholds.

6 Conclusions

In this paper we have given some formal definitions for memes, in terms of information retrieval and semantic network concepts. We have given some examples which illustrate how these definitions can be used to identify, extract and process memes from an online forum. Then we have used the meme metrics to select the memes in terms of importance and quality, for the given document set. This work lays the ground for future work in which we will process large real online forums containing free text documents (comments), and further develop the formal definitions of memes and their behaviour in different scenarios.

Acknowledgments. This research is partially supported by the Spanish MEC (HIPERGRAPH TIN2009-14560-C03-01, ARES CONSOLIDER INGENIO 2010 CSD2007-00004).

References

1. Dawkins, R. 1989. The Selfish Gene, 2nd Edition, Oxford University Press.
2. Blackmore, S.J. 1999. The Meme Machine, Oxford University Press, ISBN 019286212X.
3. Ranu, S., Chaoji, V., Rastogi, R., Bhatt, R. Recommendations to Boost Content Spread in Social Networks, in: Proc. World Wide Web 2012, April 16–20, 2012, Lyon, France, pp. 530–538.
4. Heylighen, F. and Chielens, K. 2013. Cultural Evolution and Memetics. Article prepared

for the Encyclopedia of Complexity and Systems Science, Editors: Robert A. Meyers Ph. D. ISBN: 978-0-387-75888-6 (Print) 978-0-387-30440-3 (Online). Available at: http://pespmc1.vub.ac.be/Papers/Memetics-Springer.pdf, article accessed on 15 March 2013.
5. Bordogna, G., Pasi, G. "An Approach to Identify Ememes on the Blogosphere," wi-iat, vol. 3, pp.137-141, 2012 IEEE/WIC/ACM Int. Conf. on Web Intelligence and Intelligent Agent Technology, 2012.
6. Leskovec, J., Backstrom, L. and Kleinberg, J. 2009. Meme-tracking and the dynamics of the news cycle. In Proc. 15th ACM SIGKDD Int. Conf. on Knowledge discovery and data mining (KDD '09). ACM, New York, NY, USA, 497-506.
7. Simmons, M.P., Adamic, L.A. and Adar, E. 2011. Memes Online: Extracted, Subtracted, Injected, and Recollected. In Proc. ICWSM, 2011.
8. Nettleton, D.F. 2013. Data mining of social networks represented as graphs, Computer Science Review (2013), doi:10.1016/j.cosver.2012.12.001.
9. Baydin, A.G., López de Mántaras, R. 2012. Evolution of ideas: A novel memetic algorithm based on semantic networks, Evolutionary Computation (CEC), 2012 IEEE Congress on , vol., no., pp.1,8, 10-15 June.
10. Szumlanski S. and Gomez, F. 2010. Automatically acquiring a semantic network of related concepts, Proc. 19th ACM Int. Conf. on Information and knowledge management, Oct. 26-30, 2010, Toronto, ON, Canada.
11. Miller, G.A. 1995. WordNet: A Lexical Database for English. Communications of the ACM Vol. 38, No. 11: 39-41.
12. Jiang J, Conrath D: Semantic similarity based on corpus statistics and lexical taxonomy. Proc. of Int. Conf. on Research in Computational Linguistics 1997, Taiwan, pp. 19-33.
13. Chen, Z., Gangopadhyay, A., Karabatis, G., McGuire, M., Welty, C. Semantic Integration and Knowledge Discovery for Environmental Research. Journal of Database Management, 18(1), 43-67, January-March 2007.
14. Kok, S. and Domingos, P. Extracting Semantic Networks from Text Via Relational Clustering. ECML PKDD '08 Proc. 2008 European Conf. on Machine Learning and Knowledge Discovery in Databases - Part I, pp. 624 - 639.
15. Banko, M., Cafarella, M.J., Soderland, S., Broadhead, M. and Etzioni., O. Open information extraction from the web. In Proc. IJCAI-2007, Hyderabad, India, 2007. AAAI Press.
16. Sanderson, M. and Croft, B. 1999. Deriving concept hierarchies from text. In Proc. 22nd Int. ACM SIGIR Conf. on Research and Development in Information Retrieval (SIGIR '99). ACM, New York, NY, USA, 206-213.
17. Oh, J., Kim, T., Park, S. and Yu, H. 2012. PubMed Search and Exploration with Real-Time Semantic Network Construction. Technical report, 2012. URL http://dm. postech. ac. kr/techreport/TechReport-POSTECH-2012-03-SemanticNetwork. pdf.
18. Ferre-i-Cancho, R. And Solé V., R. The Small World of Human Language, In Proceeding Real Society London B (2001) 268, pp. 2261-2265.
19. Sowa, J.F. 1991. Principles of Semantic Networks: Explorations in the Representation of Knowledge. San Mateo, Morgan Kaufmann.
20. Mihalcea, R. and Radev, D. 2012. Graphs-Based Natural Language Processing and Information Retrieval; Cambridge University Press.