Towards Semantic Evaluation of Information Retrieval

Piotr Wasilewski

Faculty of Mathematics, Informatics and Mechanics,
University of Warsaw
Banacha 2, 02-097 Warsaw, Poland
piotr@mimuw.edu.pl

Abstract. The paper discuss fundamentals of semantic evaluation of information retrieval systems. Semantic evaluation is understood in two ways. Semantic evaluation *sensu stricto* consists of automatic global methods of information retrieval evaluation which are based on knowledge representation systems. Semantic evaluation *sensu largo* includes also evaluation of retrieved results presented using new methods and comparing them to previously used which evaluated unordered set of documents or lists of ranked documents. Semantic information retrieval methods can be treated as storing meaning of words which are basic building blocks of retrieved texts. In the paper, ontologies are taken as systems which represent knowledge and meaning. Ontologies serve as a basis for semantic modeling of information needs, which are modeled as families of concepts. Semantic modeling depends also on algorithmic methods of assigning concepts to documents. Some algebraic and partially ordered set methods in semantic modeling are proposed leading to different types of semantic modeling. Then semantic value of a document is discussed, it is relativized to a family of concepts and essentially depends on the used ontology. The paper focuses on sematic relevance of documents, both binary and graded, together with semantic ranking of documents. Various types of semantic value and semantic relevance are proposed and also some semantic versions of information retrieval evaluation measures are given.

Keywords: information retrieval, semantic information retrieval, semantic search engine, semantic evaluation, evaluation methodology, information need, semantic modeling of information need, semantic relevance, semantic value of document, semantic valuation measure.

1 Introduction

The research presented in this paper is aimed at lying foundations for semantic evaluation of effectiveness information retrieval methods. Semantic evaluation is understood in two ways. Semantic evaluation *sensu stricto* consists of automatic global methods of information retrieval evaluation which are based on knowledge representation systems. Semantic evaluation *sensu largo* includes also evaluation

R. Bembenik et al. (Eds.): Intelligent Tools for Building a Scient. Info. Plat., SCI 390, pp. 107–120.
springerlink.com

of new methods of presentation of retrieved results and their comparison to previously used methods which evaluated unordered set of documents or lists of ranked documents.

In traditional evaluation methodologies human judges assessing relevance to documents on the basis of their knowledge and understanding of meaning of words appearing in judged texts. Knowledge and meaning are placed in judges' minds. Semantic evaluation methods are proposed in analogy to this situation: they are based on knowledge representation systems which are artificial stores of knowledge and meaning. In the paper, as in SYNAT project, ontologies are taken as knowledge representation systems. They can be view also as corresponding to conceptual hierarchies stored in human minds. On the theoretical basis of this correspondence, the paper introduce semantic modeling of user's information needs. Semantic modeling depends on ontologies as well as on algorithmic methods of assigning concepts to documents. Using algebraic and partially ordered set methods, various ways of semantic modeling are proposed. Semantic models of information needs serve as basis for introducing semantic value of documents. The paper proposes also semantic relevance of documents. It is a key notion of semantic evaluation of information retrieval. Semantic relevance is calculated on the basis of semantic values and is automatically assigned to retrieved texts. Semantic values of documents serves as a bridge between semantic modeling of information needs and semantic relevance of documents.

The paper has the following organization: Section 2 presents basic concepts and principles of traditional information retrieval evaluation. Section 3 discusses elements of semantic information retrieval. It presents briefly ontologies as knowledge representation systems and then introduce basic ideas of semantic evaluation of information retrieval. Section 4 discusses briefly semantic modeling of information needs and using algebraic methods proposes five types of semantic modeling. On the basis of semantic modeling, in Section 5 semantic value of documents is presented together with five types of semantic value. Section 6 discusses key concepts of semantic evaluation: semantic relevance (binary as well as graded) together with some remarks on semantic ranking of documents. Section 7 focuses on semantic evaluation of semantic information retrieval methods presenting tamped earth test (or duel test) as a way of comparing of such methods. This section if followed by Conclusions. Section 2 and first part of section 3 are reviews of a state of the art while the second part of the section 3 and the next sections introduce foundations for semantic evaluation of information retrieval.

2 Fundamentals of Information Retrieval Evaluation

Evaluation methods of information retrieval systems are aimed at reflecting how well results of searching meet user's information expectations. Currently, two broad classes of evaluation can be distinguished: system evaluation and user based evaluation (Vorhees, 2002). User based evaluation methods measure user's satisfaction with the system while system evaluation methods measure how well the system can rank documents (Vorhees, 2002).

Key notions in information retrieval evaluation are *information need* and *relevance of a document*. From the very beginning of the notion of information need, it was highlighted that information need has both conscious and unconscious components: it is a desire of an individual person or group of people to find and get information satisfying their conscious or unconscious needs (demands) (Taylor, 1967). In other words, IN is a topic on which a user would like to know more, and it is distinguished from the query - a data structure which is entered to a IR system by a user in order to communicate information need (Manning et al., 2008). *Relevance* indicates how well a document or set of documents satisfies the user's information needs (Cuadra and Katter, 1967). In other words, the document is *relevant* if it is perceived by the user as containing valuable information with regard to its information needs (Manning et al., 2008). Relevance is traditionally of binary nature: the document is relevant or irrelevant (Butcher et al., 2010; Manning et al., 2008), the vast majority of test collections assume this, however, in the first Cranfield experiments a five-point scale of relevance was used (Cleverdon, 1967; Vorhees, 2002; Voorhees and Harman, eds., 2005). Recently, *graded relevance* again become used in the evaluation experiments (Najork et al., 2007; Butcher et al., 2010).

For the standard way of ad hoc measuring the effectiveness of information retrieval systems test collections consisting of three components are used (Vorhees, 2002; Manning et al., 2008):

- A set of documents,
- The test kit of information needs, expressed as queries,
- A set of relevance propositions, usually binary assignments of labels *relevant* or *irrelevant* to each pair consisting of query and document.

Historically, the first proposed paradigm of this type was the Cranfield paradigm (Cleverdon, 1967; Vorhees, 2002; Voorhees and Harman, eds., 2005; Manning et al., 2008). Currently, more modern versions of this paradigm are used together with bigger test collections. They are discussed and developed at few conferences: the TREC conference (Text Retrieval Conference), organized in the U.S. and at two conferences dedicated to inter-language information retrieval NTCIR (NII Test Collections for IR Systems), which focuses on East-Asian languages and CLEF (Cross Language Evaluation Forum), which focuses on European languages (Manning et al., 2008).

This traditional way of information retrieval evaluation is based on two fundamental assumptions (Butcher et al., 2010):

- Having a given user's information need, represented by a query, each document in a given set of texts is relevant or irrelevant with regard to this information need.
- The relevance of the document d depends solely on the information need and the d itself, being independent from ranking of other documents in the collection by a search engine.

Methods of evaluating unordered sets of documents were historically first between information retrieval evaluation methods (van Rijsbergen, 1979). Then

methods evaluating ranked retrieved sets of documents appeared. The former include such classic measures as recall, precision, specificity, fallout, the latter include interpolated precision, mean average precision, reciprocal rank cumulative gain measures (Manning et al., 2008; Butcher et al., 2010).

3 Semantic Information Retrieval

Semantic retrieval is a new type of information retrieval. The semantic search engine, prepared under project SYNAT, will be one of the first systems of that type. The evaluation methods proposed up to now, are related to information retrieval systems, which can be described as linguistic/syntactic. In such systems searching is based on the presence of words in documents. In the semantic information retrieval the meaning of words are involved, whereas searching is done by looking at the knowledge contained in documents. Thus, semantic information retrieval must be based on the some way of knowledge representation. In the project SYNAT, for the purpose of knowledge representation ontologies are selected, they are presented as sets of concepts connected by various relations, mainly by the relation of subsumption (is-a relation), however being-a-part-of relation or other relations are also admissible (Breitman et al., 2007; Buitelaar and Cimi, eds., 2007; Colomb, 2007; Staab and Studer, eds., 2009). Additionally, we treat concepts from ontologies as meanings of words while the knowledge in ontologies is contained in relations, or also in the concepts, assuming that they are defined on the basis of attributes/slots[1]. Therefore, methods for assessing the effectiveness of semantic information retrieval and semantic relevance of the documents should be based on ontologies. By this conclusion, we break the second assumption of the traditional information retrieval evaluation pointed out in Section 2:

> semantic relevance of document d depends not only on information need α and the document itself but also on the ontology O_1: when O_1 is changed to another ontology O_2, document d in the context of query α may get a different semantic relevance.

Information retrieval systems are typical examples of human - computer interaction systems. In any information retrieval system, four elements can be distinguished:

- a user's mind[2] being a source of information needs and formulated queries,
- user's interface used for entering queries
- search engine operating with an inverted index and retrieving documents,
- data repository, storing all collected documents

[1] In taxonomies, being ontologies of the simplest form, one cannot claim that taxonomic topics contain knowledge, in this case knowledge is only contained in the subsumption relation.

[2] Understood following cognitive science as an information processing system.

In the semantic information retrieval system, a module of semantic searching is equipped with knowledge representation system, e.g. with a given ontology, while meanings are assigned to the words from document on the basis of this ontology. Therefore, in the semantic information retrieval system, meaning and further, knowledge are located in two modules of the system: in the user's mind in which they are components of information need and in the ontology incorporated in the system. In the SYNAT project it is also planned to develop user's interface to a dialogue model for user - search engine interactions, which will conduct a dialogue with the user aimed at specification of a query and driving the searching of documents or presentation of retrieved results. An important function of the module will be the translation of a query entered by the user and expressed in natural language, onto a query in an ontology based a descriptive logic language. In this translation, the ontology from a semantic search engine will be also involved.

Note that in the context of ontologies adopted in the project for knowledge representation, semantic evaluation of information retrieval should be distinguished from the evaluation of semantic information retrieval with respect to knowledge. Having two different semantic search engines W_1 and W_2 and basing them on the same sematic component, e.g., on the same ontology[3], semantic search factor is controlled, therefore we can evaluate effectiveness of search engines W_1 and W_2 using classical nonsemantic methods. We can use such global methods even when two different search engines are supported by two different ontologies, but in this case it cannot be certain whether e.g. indexing algorithms or ontologies are responsible for the effectiveness of retrieval. Similarly, it is possible that having two nonsemantic search engines, we can evaluate their effectiveness using a semantic method of effectiveness evaluation based on given ontology O_1.

It is worth noting that notions of evaluation of semantic information retrieval and semantic evaluation of information retrieval are independent, i.e. all four possibilities can hold. To two possibilities pointed above, one has to add the possibility practiced so far, i.e. nonsemantic evaluation of nonsemantic information retrieval and the forthcoming possibility of semantic evaluation of semantic information retrieval.

4 Semantic Modeling of Information Needs

An information need arising in the user's mind consist, inter alia, of concepts. However, an information retrieval system has no access to the conceptual frames in the user's mind. Communication between the mind and the information retrieval system is done through a query formulated and entered by the user expressing his/her information need. A query is a data structure usually consisting of words. In the sequel, we assume that words contained in the query and referring to concepts (terms) are mapped by the system to concepts included in the ontology of the system. Let us note that this mapping is in fact assigning to a

[3] Semantic search is supported by the same ontology.

query its meaning in a given ontology and that the context of this ontology is essential: the same query in two different ontologies can have two different meanings. This reveals the nature of semantic modeling of queries: a given query is semantically modeled by a family of concepts interpreted as meanings of words contained in the query.

In the following considerations, we adopt simplifying assumption about the ontology: by an ontology we mean set of concepts O_1 partially ordered by subsumption relation $\leqslant$: $\langle O_1, \leqslant \rangle$. If it will not lead to confusion (a subsumption relation will be understood from the context), to ontology $\langle O_1, \leqslant \rangle$, as a partial order, we will refer also by O_1.

Let $\langle O_1, \leqslant \rangle$ be an ontology used by a given semantic search engine. Hereafter we model the information need semantically as a set of concepts from ontology O_1: $\{C_1, ..., C_n\} \subseteq O_1$ determined in some way by query q expressing this information need. Such family we will call *a semantic model of query* q or *a semantic model of information need*. Because information need is always expressed in the form of a query, we will also briefly say that the family of concepts models semantically the query.

1. The simplest way of semantic modeling of the information need expressed by query q is to take concepts from the ontology: O_1 which are assigned by the system to terms contained in query q. Such family of concepts we will denote by $O_1(q)$.
2. Another way of modeling query q is to take additionally concepts from ontology O_1 which are placed between concepts from family $O_1(q)$ which are comparable with respect to subsumption relation $\leqslant$. Such family we will denote by $O_1[q]$, in other words:

$$O_1[q] := \{D | \exists A, B \in O_1(q); A \leqslant D \leqslant B\} \tag{1}$$

   Let us note that family $O_1[q]$ can be empty even when $O_1(q)$ is nonempty, and this is when $O_1(q)$ is an anti-chain, i.e. any two concepts from $O_1(q)$ are not comparable with respect to subsumption relation $\leqslant$. Taking into account such possibility, we can introduce next ways of sematic modeling of information needs.
3. Family $O_1(q)$ can be taken as a set of generators of a complete lattice: we take family $O_1(q) \subseteq O_1$ as partially ordered set $\langle O_1(q), \leqslant_{|\langle O_1(q)} \rangle$ and then we take the Dedekind-MacNeille completion of $\langle O_1(q), \leqslant_{|\langle O_1(q)} \rangle$ which is a complete lattice[4] For the family semantically modeling query q we take the universe of this lattice denoted by $L[O_1(q)]$.
4. Let us note that family $L[O_1(q)]$ not necessarily contains e.g. all upper bounds of family $O_1(q)$ in set $\langle O_1, \leqslant \rangle$ (upper bounds of family $O_1(q)$ are superconcepts of all concepts from family $O_1(q)$)[5]. In order to consider all

[4] On of the methods of construction of the Dedekind-MacNeille completion is creating a concept lattice (Wille, 1982; Ganter and Wille, 1999) for a given partially ordered set (see Dedekind completion theorem in Ganter and Wille, 1999). Creating finite concept lattices has a computational character.

[5] All lower bounds of family.

elements somehow generated from family $O_1(q)$ we can proceed in two ways. Firstly, the Dedekind-MacNeille completion of whole ontology O_1 is taken, denote the universe of this lattice by $L[O_1]$ (note that $O_1 \subseteq L[O_1]$. Then take family $O_1(q)$ as a set of complete generators and generate complete sublattice $Sg_{L[O_1]}(O_1(q))$ of the complete lattice $L[O_1]$. For the family semantically modeling query q we take family $Sg_{L[O_1]}(O_1(q))$.

5. Secondly, take the family of all concepts from ontology O_1 which ar comparable by the subsumption relation with at least one concept from family O_1, i.e. take the family of the form:

$$FI_{O_1(q)} = \bigcup_{A \in O_1} (A] \cup \bigcup_{A \in O_1} [A), \tag{2}$$

where $(A]$ and $[A)$ are respectively a principal filter and a principal ideal determined by concept A in partially ordered set $\langle O_1, \leqslant \rangle$. Then take the Dedekind-MacNeille completion of partially ordered set $\langle FI_{O_1(q)}, \leqslant_{|FI_{O_1(q)}} \rangle$, the universe of this complete lattice will be denoted by $L[FI_{O_1(q)}]$. For the family semantically modeling query q we take family $L[FI_{O_1(q)}]$.

Note that two last methods of modeling of information needs outlined above are different and have their own advantages and disadvantages. Lattices $Sg_{L[O_1]}(O_1(q))$ and $L[FI_{O_1}]$ do not have to be isomorphic. The first method is computationally expensive because it requires construction of lattice $L[O_1]$, however, only a half of the job should be done, since every ontology, as a tree, is a complete semi-lattice. In the second method, some elements of lattice $L[FI_{O_1}]$ do not have to belong to $L[O_1]$, thus they do not have to be related to concepts from ontology O_1. On the other hand, the question of computational complexity of the first method is significant only in the case of dynamically changing ontology O_1 requiring to online computation of lattice $L[O_1]$ whereas elements from lattice $L[O_1]$ unrelated to concepts from ontology O_1 in some contexts may be regarded as an advantage rather than disadvantage, for example if such concept will appear in some documents this can be seen as a reason for adding it to ontology O_1. It is worthy to note also that the above list is open and other methods of semantic modeling of information needs can be proposed.

Finally, semantic modeling of information needs can be used for constructing semantic information retrieval methods as well as semantic evaluation measures of retrieval effectiveness. In this paper we investigate the latter possibility but definitely the former is also worth of exploration.

5 Semantic Value of Documents

On the basis of semantic modeling of queries now we can move now to semantic characterization of documents. First notion of this kind is a semantic value of a document.

Having given family of concepts $\Phi = \{C_1, ..., C_j$ of a given ontology O_1 we can determine a semantic value of document d. Let $C_\Phi(d, C_i) = 1$ for $1 \leqslant i \leqslant j$,

if C_i is contained in document d, otherwise $C_\Phi(d, C_i) = 0$. Semantic value of document d with respect to family Φ has the following form:

$$\sum_{i=1}^{k} C_\Phi(d, C_i). \tag{3}$$

Firstly, note that a sematic value can be calculated for a set of documents. In such case also an average sematic value of a set of documents can be calculated. Note also that family Φ does not have to be interpreted as a semantical model of an information need/querry. For example, as family Φ can be taken the whole ontology, in this case a semantic value of documents can be used for characterization of this ontology on the basis of a given set of documents or a a basis for comparison of two different ontologies. It is also possible that semantic value is not an integer, it can hold in the case when an algorithm of mapping concepts to documents will describe particular concepts in the context of a given document by numbers other than 0 and 1. In the sequel, all consideration will admit this possibility.

Let us note that for family Φ different families of concepts can be taken representing different methods of semantic modeling of information needs, including the five methods outlined above. And so, keeping the way of enumerating of this list we get the following types of a sematic value of a given document d with respect to ontology O_1:

$$SV_1(d) = \sum_{i=1}^{k} C_{O_1(q)}(d, C_i), \tag{4}$$

$$SV_2(d) = \sum_{i=1}^{k} C_{O_1[q]}(d, C_i), \tag{5}$$

$$SV_3(d) = \sum_{i=1}^{k} C_{L[O_1(q)]}(d, C_i), \tag{6}$$

$$SV_4(d) = \sum_{i=1}^{k} C_{Sg_{L[O_1]}(O_1(q))}(d, C_i), \tag{7}$$

$$SV_5(d) = \sum_{i=1}^{k} C_{L[FI_{O_1(q)}]}(d, C_i). \tag{8}$$

Note also that family Φ can represent semantical modeling of many information needs at the same time, being simply set theoretical union of sematic models of particular information needs.

It is worthy to note that sematic value of documents essentially depends on an ontology taken as a basis for modeling of documents. Particularly, it is reflected by measures of sematic value from SV_2 to SV_5. Consider now the fact that semantic value can be used to determine the semantic relevance of documents as

well as semantic ranking of documents. The first shows its usefulness in semantic evaluation of information retrieval, while the second can be applied both in semantic evaluation and semantic information retrieval.

6 Semantic Relevance and Semantic Ranking of Documents

A document is relevant if it is perceived by a user as containing valuable information with respect to of his/her personal information needs (Cuadra and Katter, 1967; Manning et al. 2008). Relevance indicates how well a document or set of documents satisfies the user's information needs (Cuadra and Katter, 1967). In order to be able to talk about semantic relevance , there must be some connection between the evaluation of semantic relevance of a document and a given information need. Semantic value of the document seems to give the basis for such a relationship, because through the semantic modeling it bounds information needs of users with the documents.

6.1 Binary Semantic Relevance

Having given document d, query q, family of concepts $\Phi = \{C_1, ..., C_j$ and based on Φ a semantic value of d we can determine binary a *sematic relevance of d with respect of q* in the following way:

$$\text{if } \sum_{i=1}^{k} C_{\Phi}(d, C_i) = k, \text{ then } d \text{ is semantically relevant with respect to } q,$$

$$\text{if } \sum_{i=1}^{k} C_{\Phi}(d, C_i) < k, \text{ then } d \text{ is semantically irrelevant with respect to } q \\ (d \text{ is not semantically relevant w.r.t. } q).$$

In other words, the document d is relevant with respect to query q on the basis of family of concepts Φ, when every of the concepts from family Φ is contained in document d. Note that in conjunction with the five types of semantic value outlined above, we have at least five types of binary semantic relevance. It should be noted that, as in the case of methods of semantic modeling of information needs or types of semantic value of documents, a list of types of semantic relevance is open. It is also worth noting that the large cardinalities of family Φ of such approach to binary semantic relevance can be very restrictive.

Having the five types of binary semantic relevance, it is worth noting that we also have five semantic versions of each of the classical measures of the effectiveness of informational retrieval based on relevance of documents binary understood. For example, we consider the semantic version of the average precision (where $\Phi = \{C_1, ..., C_j$ is a family of concepts which is a semantic model of a given information need):

$$AP^{\Phi} = \frac{1}{|Rel_{\Phi}|} \cdot \sum_{i=k}^{|Res|} relev(k) \cdot Pr@k_{\Phi} = \frac{1}{|Rel_{\Phi}|} \cdot \sum_{i=k}^{|Res|} relev(k) \cdot \frac{|Res[1,...,k] \cap Rel_{\Phi}|}{k}, \tag{9}$$

where Rel_{Φ} is a set of documents relevant with respect to a given information need on the basis of family of concepts Φ being the semantic modeling of this information need, Res is a set of all retrieved documents and $Res[1,...,k]$ consists of the top k documents ranked by the system, while $relev(k) = 1$ if k-document in Res is relevant, $relev(k) = 0$ otherwise.

Other example is semantic version of geometric mean average precision for n information needs:

$$GMAP(AP_1^{\Phi},...,AP_n^{\Phi}) = \sqrt[n]{\prod_{i=1}^{n}(AP_i^{\Phi} + \varepsilon)} - \varepsilon, \tag{10}$$

where ε is a constant aimed at eliminating pathologies when one of the average semantic precisions, AP_i^{Φ}, is equal to 0.

6.2 Graded Semantic Relevance

The simplest way of introducing graded semantic relevance for document d, query q and family of concepts $\Phi = \{C_1,...,C_n\}$ semantically modeling query q is to identify semantic relevance with semantic value. Other way is to normalize graded semantic relevance to the unit $[0,1]$:

$$IS_{\Phi}(d,q) = \frac{\sum_{i=1}^{n} C_{\Phi}(d,C_i)}{n}, \tag{11}$$

where $IS_{\Phi}(d,q)$ denotes semantic relevance of document d with respect to query $q$on the basis of family of concepts Φ. Notre that $IS_{\Phi}(d,q) = 1$ if, and only if document d is semantically relevant with respect to query q by means of binary semantic relevance. A value of normalized semantic relevance can be average to given set of queries Q:

$$IS_{\Phi}(d) = \frac{1}{|Q|} \cdot \sum_{q \in Q} IS_{\Phi}(d,q). \tag{12}$$

Measure IS_{Φ} (for queries as well as sets of queries) we will call generally normalized semantic relevance.

For example, we present semantic versions of normalized discounted cumulative gain measure for graded relevance on the basis of family of concepts Φ: $nDCG_{\Phi}$. Let be given a list of ranked documents of which every has assigned normalized semantic relevance. For this list we create the semantic gain vector G_{Φ} composed of normalized semantic relevance of documents from the list, a value of relevance placed on i place of the vector G_{Φ} we denote by $G_{\Phi}[i]$. Therefore, $G_{\Phi}[i] = IS_{\Phi}(d_i)$. Then we calculate the cumulative semantic gain vector

CG_Φ of which value of k element is a sum of values of elements of the vector G_Φ from 1 to k:

$$CG_\Phi[k] = \sum_{i=1}^{k} G_\Phi[i]. \tag{13}$$

Then we calculate discounted semantic gain:

$$DCG_\Phi[k] = \sum_{i=1}^{k} \frac{G_\Phi[i]}{\log_2(1+i)}. \tag{14}$$

Then the perfect semantic gain vector G'_Φ is constructed, it consists of elements of the semantic gain vector G_Φ, where if $i \leqslant j$, then $G'_\Phi[i] \geqslant G'_\Phi[j]$. Then the cumulative perfect semantic gain vector CG'_Φ and the discounted cumulative perfect semantic gain vector DCG'_Φ are calculated. The last step is a normalization of the discounted cumulative semantic gain vector by the discounted cumulative perfect semantic gain vector:

$$nDCG_\Phi[k] = \frac{DCG_\Phi[k]}{DCG'_\Phi[k]}. \tag{15}$$

6.3 Semantic Ranking of Documents

Documents can be semantically ranked, e.g., by means of their semantic value. In this case we are dealing with at least five types of semantic ranking of documents. Note that the semantic ranking of documents can be naturally combined with the graded semantic relevance, e.g., with the normalized semantic relevance. Let us note that the combination of semantic ranking based on the semantic value and the normalized semantic relevance we always get the ideal semantic gain vector.

7 Tamped Earth Test or Duel of Two Search Engines

It has to be underlined that both semantic information retrieval and semantic evaluation of information retrieval (semantic as well as nonsemantic) significantly depend both on the algorithms for indexing documents and on concepts from ontologies (adopted in SYNAT project as a way of knowledge representation). This can be seen as disadvantage in the context of constructing semantic evaluation measures: such measures can prefer search engines using the same conceptual indexing algorithms. This problem can be partially solved in the future by establishing conventionally some standardized conceptual indexing algorithm/algorithms. However, it is still only a partial solution. Before further discussion, let us accept a notational convention:

> retrieved results of search engine W_i for set of documents D_j we will denote by $W_i(D_j)$. A semantic evaluation measure M which is based on conceptual indexing algorithm p_s and ontology O_t will be denoted by $M(p_s, O_t)$. Search

engines with built in algorithms and/or using some ontologies we will denote multiplicatively, for example $W_i a_k O_t$ denotes search engine W_i with built in algorithm a_k and using ontology O_t, while $W_i a_k O_t(D)$ denotes its retrieved results on set of documents D.

Now, take into account the worst case from the perspective of the pointed above disadvantage. Assume that we have two search engines W_1 and W_2 which use two different conceptual indexing algorithms, respectively, p_1 and p_2, and two different ontologies O_1 and O_2. In such case, one can evaluate their retrieval results for a given set of documents by means of a semantic measure M using algorithms p_1 and p_2 and ontologies O_1 and O_2 in all possible arrangements, i.e. four versions of the appropriate measure of M: $M(p_1, O_1)$, $M(p_1, O_2)$, $M(p_2, O_1)$, $M(p_2, O_2)$. Note that each of these semantic versions of measure M is unjust and therefore, as a fair evaluation, the average value of all four measures can be taken. Note also that the following result would be particularly striking: namely, if one search engine defeat a second search engine using the opponent's weapon (the algorithm and ontology), for example, if search result $W_1(D)$ was better than $W_2(D)$ with respect to evaluation measure $M(p_2, O_2)$. In this case, W_1 would be particularly convincing winner od a duel - hence the name of the test. Also a nonsemantic version of the test is possible: evaluation of retrieved results of various search engines, e.g. $W_1 p_2 O_2$ vs $W_2 p_2 O_2$, by means of classical nonsemantic measures.

Let us also note that the tamped earth test has four types of variables:

- search engines,
- indexing algorithms,
- ontologies,
- evaluation measures.

Therefore, fixing of particular variables gives methods of testing of other elements. For example, fixing a search engine, indexing algorithm and evaluation measure, one can test various ontologies, e.g. O_1 and O_2 comparing $W_1 p_1 O_1(D)$ vs $W_1 p_1 O_2(D)$ with respect to appropriate versions of evaluation measure M: $M(p_1, O_1)$ and $M(p_1, O_2)$.

8 Conclusions

Semantic modeling of information needs and semantic values of documents introduced in this paper can be applied in semantic evaluation methods as well as in constructing new semantic retrieval methods.

To be applicable, methodology of evaluating the effectiveness of information retrieval must include the following elements (Butcher et al., 2010):

- characterization of the intended purpose of information retrieval method,
- measure, which quantitatively shows how well this goal is satisfied,
- precise, accurate and economical measurement technique,
- estimation of measurement error.

The estimation of measurement error is a problem solvable in a standard way for all evaluating methodologies including semantic ones (van Rijsbergen, 1979; Butcher et al., 2010). Therefore, research should be focused on the first three topics.

Some exemplary semantic evaluation measures are given in this paper, however, next semantic evaluation measures should be proposed in the future. Since methods of assessing of semantic relevance presented in this paper are automatic, it meets the third point. Implementation and testing of these methods will reveal their usefulness. They can be also compared to the traditional evaluation methods based on assessing made by human judges. Such comparison can show how far away semantic evaluation methods depart from them. Satisfaction of the first point involves a combination of further theoretical research with computational simulations of the proposed methods of semantic evaluations of information retrieval and this will be done in the future. Especially interesting are investigations into semantic evaluation of effectiveness of semantic information retrieval.

Let us note that generally semantic evaluation methods are not necessarily alternative to human judging methods. They can be used also to support assessing process made by human judges as it is done analogically in the case of manual indexing of documents from PubMed search engine, where an automated indexer indexes the title and abstract and supports the manual indexer by providing a list of potential MeSH ontology keywords (Berry and Browne, 2005).

Acknowledgements. Author would like to thank Andrzej Skowron, Hung Son Nguyen, Dominik Ślęzak, Wojciech Jaworski, Wojciech Świeboda and other colleagues from the SYNAT group for their critical comments and valuable discussions helping to improve the paper. Research was supported by the grant N N516 077837 from the Ministry of Science and Higher Education of the Republic of Poland and by the National Centre for Research and Development (NCBiR) under the grant SP/I/1/77065/10 by the Strategic scientific research and experimental development program: "Interdisciplinary System for Interactive Scientific and Scientific-Technical Information".

References

1. Berry, M.W., Browne, M.: Understanding Search Engines. Society for Industrial and Applied Mathematics (2005)
2. Breitman, K.K., Casanoca, M.A., Truszkowski, W.: Semantic Web: Concepts, Technologies and Applications. Springer, Heidelberg (2007)
3. Buitelaar, P., Ciminao, P. (eds.): Ontology Learning and Population: Bridging the gap between Text and Knowledge. IOS Press (2008)
4. Butcher, S., Clarke, C.L.A., Cormack, G.V.: Information Retrieval: Implementing and Evaluating Search Engines. Massachusetts Institute of Technology Press (2010)
5. Cleverdon, C.W.: The Cranfield tests on index language devices. In: Aslib Proceedings, vol. 19, pp. 173–192 (1967); Reprinted in Readings in Information Retrieval, Sparck-Jones, K., Willett, P. (eds.) Morgan Kaufmann (1997)
6. Colomb, R.M.: Ontology and the Semantic Web. IOS Press (2007)

7. Cuadra, C.A., Katter, R.V.: Opening the black box of relevance. Journal of Documentation 231(4), 291–303 (1967)
8. Ganter, B., Wille, R.: Formal Concept Analysis: Mathematical Foundations. Springer, Heidelberg (1999)
9. Jarvelin, K., Kekalainen, J.: Cumulated gain-based evaluation of IR techniques. ACM Transactions on Information Systems 20(4), 422–446 (2002)
10. Manning, C.D., Raghavan, P., Schutze, H.: Intorduction to Information Retrieval. Cambridge University Press (2008)
11. Najork, M.A., Zaragoza, H., Taylor, M.J.: HITS on the Web: How does it compare. In: Proceedings of the 30th Annual International ACM SIGIR Conference on Research and Development in Information Retrieval, pp. 471–478 (2007)
12. Robertson, S.: On GMAP and other transformations. In: Proceedings of the 15th ACM International Conference on Information and Knowledge Management, pp. 78–83 (2006)
13. Staab, S., Studer, R. (eds.): Handbook on Ontologies. Springer, Heidelberg (2009)
14. Taylor, R.S.: Process of Asking Questions. American Documentation 13, 291–303 (1967)
15. van Rijsbergen, C.J.: Information Retrieval, 2nd edn. ch. 7. Butterworths (1979)
16. Voorhees, E.M.: The philosophy of information retrieval evaluation. In: Peters, C., Braschler, M., Gonzalo, J., Kluck, M. (eds.) CLEF 2001. LNCS, vol. 2406, pp. 355–370. Springer, Heidelberg (2002)
17. Voorhees, E.M., Harman, D.K. (eds.): TREC. Experiment and Evaluation in Information Retrieval. Massachusetts Institute of Technology Press (2005)
18. Wille, R.: Restructuring lattice theory. In: Rival, I. (ed.) Ordered Sets. Reidel (1982)