An Information Retrieval Models taxonomy based on an Analogy between cognitive science and Information Retrieval

Yaël Champclaux (*, **), **Taoufiq Dkaki** (*, **), **Josiane Mothe** (*, **)
champ@irit.fr, dkaki@irit.fr, mothe@irit.fr
(*) IRIT
(**) Université Paul Sabatier, Toulouse, France

Mots clefs :
Taxonomie, comparaison, recherche d'information, science cognitive, modèles de recherche d'information, modèles de jugement de similarité.
Keywords:
Taxonomy, comparison, Information retrieval, cognitive science, information retrieval models, similarity judgment models.
Palabras clave:
Taxonomía, comparación, recuperación de Información, Ciencia Cognitiva, modelos de recuperación de información, modelos de criterio de similitud.

ABSTRACT:

The task of ad hoc information retrieval (IR) consists in finding documents in a corpus that are relevant to an information need specified by a user's query. In a retrieval model which is an abstraction on the IR process, there are two fundamental aspects: how documents and queries are represented – and stored-, and how they are compared in order to decide which document to retrieve to the user. IR model taxonomies permit to have a view on existing models, help to understand the relationship that exists among them by giving a view on general and specific properties of a given model. Several IR model taxonomies have been proposed, they are generally based on the mathematical aspects of the model. We choose a different approach to propose a new taxonomy of IR models based on the philosophy used to compare query and documents. Our IR model taxonomy was inspired by an analogy between automatic comparison methods in Information Retrieval Systems (IRS) and object comparison method in cognitive science. We look at the comparison problem with a cognitive point of view: comparison between entities may depend on the features of compared entities – indexing terms -, sometimes on relations between these features and sometimes on relations between entities – corpus structure -. Nowadays taxonomies of IRS do not take into account the structural aspect which becomes more and more important, that's why we propose a cognitive oriented taxonomy which highlights structural aspect.

1. Introduction

Numerous scientific disciplines influence information Science:

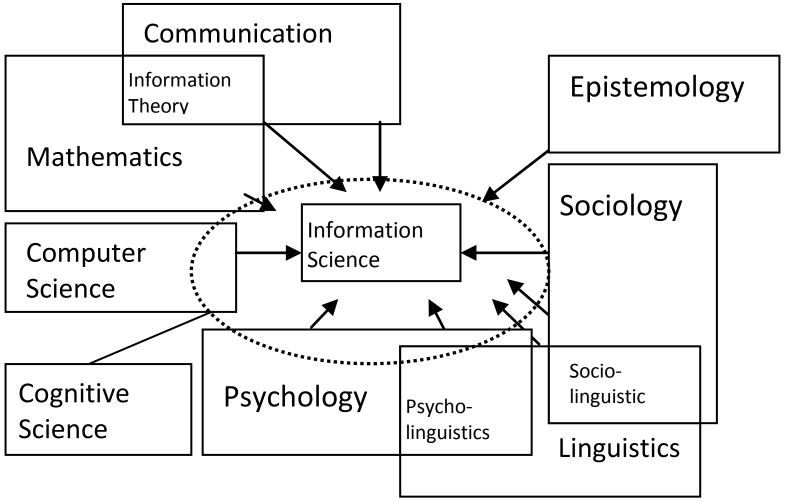

Figure 1: Information science a discipline crossroads [32]

By working on traditional IR, we found a bridge between automatic comparison in Information Retrieval Systems (IRS) and comparison between entities in cognitive science.

The main task of ad hoc information retrieval consists in finding documents in a corpus that are relevant to an information need specified by a user's query. This research problem has lead to plenty of works during the last fifty years and is far from being solved. We are interested in the most traditional subfield of IR: text retrieval. To realize the IR task, IRS process documents and queries to obtain a representation that permits their comparison: the indexing process. During this phase, words of documents and queries are usually filtered using a stop list in order to remove semantically vacuous words and select informative words, remaining words are stemmed to reduce morphological variants. The indexing process leads to obtain representations of documents and user's need. The core of an IRS is the comparison engine where documents are compared against queries to determine similar-to-query documents. The indexing process and comparison process are defined in the IR model.

We focus our interest on the comparison aspect and on how similarity is defined between documents. To do so, we choose to look at a more general problem: comparison between entities. Different comparison models have been proposed in cognitive science to formalize the comparison activity and to study criterions used to determine similarity between objects. Those different models are grouped among four families: geometric models, features based models, structural models and transformational models. We are especially interested in the fact that structural approach has to constitutes a separate model family because structural approach is becoming more and more important in the IR field and existing taxonomy do not highlight this aspect. The structure here is not related neither to structured documents nor to the INEX sense of structure – sections, paragraphs, sentence. We talk about the document corpus structure. A document is not only constituted by the indexing terms, but also by the relations induced by theses terms with other document of the collection.

Traditional taxonomies are based on the mathematical aspects of the model; we prefer to consider the mathematics as a tool to realize the model, and to classify models on the basis of the underlying philosophy.

In the reminder of this paper, we will introduce the related works on IR taxonomies, then two pre-requisite where we will present different retrieval model, and different comparison models in cognitive science that inspired our taxonomy of IR models. The proposed taxonomy will be presented in section 5.

2. Related works

Several taxonomies of IRS have been proposed [1], [2], [3], [4], and more recently [5][6], these works aim to classify research approaches in the domain of IR according to several features.

Each taxonomy chooses particular point of view to classify IRS. For example, Jian-Yun Nie proposed in [3] an IR model classification based on the type of comparison (strict or flexible) that can be done between documents and queries and on the representation level of data. Paijmans in [4] proposes to construct taxonomy of IR models to identify a generic model (the vector document model) which is used as a basis to obtain more specific models. The classification proposed in [5] consists of super-imposing two views: the first view is a vertical taxonomy that classifies IR models with respect to the way documents and queries are represented and also with respect to the framework used to determine similarity between those representations. The second view consists in a horizontal taxonomy that classifies IR objects with respect to their tasks, form –service or tool-, and context. The classification proposed in [6] that can be found on Wikipedia[1] propose to differentiate models based on two dimensions: the first dimension is the mathematical basis of models; the second includes properties of models about term interdependencies. In spoken and written language, terms are not always independent. As we said before, for the information retrieval to be efficient, documents are typically transformed into a suitable representation during indexing process. During this process, models take assumptions concerning term interdependency. In [6] three levels of interdependencies are defined: no interdependency, immanent interdependency and transcendent interdependency. No interdependency indicates that terms are considered as independent, with immanent interdependencies terms interdependencies are defined by the model itself and with transcendent interdependencies the term interdependencies is given by an external source. The classification proposed in [3] has the advantage to accurately handle models with a semantic dimension and models independent from data representation like the logical model [7]. The classification proposed in [5] has the advantage to take into account different features (representation and reasoning) that permits to classify logical models and also learning based models. Furthermore, this classification offers the opportunity to classify retrieval objects –algorithms, methods, technologies and tools- and show the relationship that exists between a retrieval model and a retrieval object. For example, vector space model is a retrieval technique that can be used for building information filtering and document clustering tools [5]. Different specific taxonomies exist: some permit to focus on non-classical models like [3], some try to propose taxonomy that permits to classify almost each retrieval model and tools related [5].

A retrieval model is an abstraction of the retrieval process. As we said before, the core issue of IRS is comparison. It is indeed an important point to model how documents and queries are to be compared. Comparison is a theme studied in cognitive science because it is fundamental in numerous cognitive tasks –

[1]2010, http://en.wikipedia.org/wiki/Information_retrieval

discovering, learning problem solving-. Searchers try to understand how comparison is achieved in order to formalize this activity and propose judgments models of similarities to predict Human judgments. Searchers place the understanding of similarity as fundamental for entities comparison.

We think that similarity is fundamental too in IR and we wonder if there is some similarities between comparison in cognitive science and comparison in IR.

We want to propose IR model taxonomy to classify the retrieval models presented in next section. To do so, we choose to exploit cognitive science works on comparison which is the subject of the section 4.

3. IR models

The Core of an IRS is the retrieval model. A model is an abstraction of the retrieval process: models have to determine how documents and queries are represented, and how similarity between them is defined. In this study, we will focus on models which use indexing process to store data (documents and queries), and different theories to compare documents representation against queries representation.

- Boolean Model (1950) is based on Boole Algebra, query and documents are considered as a set of terms; a document is relevant for a query if it contains query's terms. The Boolean model is an exact matching model. A document is either relevant or not for a given query [29].
- Vector Space Model (1970) is based on Vector space theory, query and documents are considered as vectors of terms; a document is relevant for a query if the dot product (or Cosine, or Dice coefficient, etc.) is not zero [35].
- Probabilistic Model (1976) is based on probabilistic assumptions; query and documents are considered as a set of event. An event denotes term presence or term absence; the relevance of a document for a query depends on the ratio of the conditional probability of the document given relevance occurs to the conditional probability of the document given non-relevance occurs [33].
- Extended Boolean Model (1983) introduce term weighting in Boolean retrieval. This model use vector representation and distance calculation between vectors to determine the relevance of a document versus a Boolean query [34].
- Fuzzy sets (1984) permits to introduce a gradual similarity between documents and queries in set theories based models [9].
- Connectionist or neural Model (1989) permits to represent dependencies between term and also a dynamic aspect of document representation by using neural theory. The query is the initial stimuli of the neural network. The network response is the set of documents that are activated by the initial stimuli. The level of the document node activation is used as relevance criteria [10].
- Inferential Model (1992) is used in IR to represent term interdependencies and document-term dependencies. In this model, the relevance of a document for a given query corresponds to the belief degree that the document satisfies the user's need [11].
- Latent Semantic Indexing Model (1994) is based on a mathematical operation called singular value decomposition that transform the initial document vector space into another document vector space were similar documents are closer to each other [12].
- Language Model (1998) assumes that the relevance of a document for a given query is the probability that the document generates the query [13].
- Structural Language Model (2006) incorporates the corpus-structure information — modeled using clusters of similar documents — into the language modeling framework [14].
- Graph Model (2009) exploits the document-term graph structure to determine similarity between a query and a document: the similarity not only depends on the shared indexing terms, but also on the neighboring of the query and the document [8].

Structural approaches have been used in a subfield of IR : re-ranking [15][30][31].The idea is to re-rank the top retrieved documents given by an IRS using a structural strategy i.e. extract potentially useful information from the corpus structure.

Importance of structure has been shown in web retrieval: Jeromy Carriere and Rick Kazman propose in [16] a method to combine link analysis and user's query. Sergey Brin and Lawrence Page use hyperlink structure of the web to determine the similarity between web pages [17]. In a similar approach, Kleinberg also use the web structure to determine the hub score and the authority score of a given page [18].

Cognitive science studies explore the role of relations and attributes in similarity. An attribute is an object description; a relation is a predicate that links two or more object, attributes (object description) or relations. In the next section, we will present similarity models in cognitive science.

4. Similarity Judgments Models

Performing comparison through the assessing of similarity is fundamental for cognition; many activities require the use of comparison and consequently of similarity: problem solving – similar problem may have similar solutions -, categorization, and anticipation – similar situation may have the same issue. Many studies tried to circumscribe the criterions used to define if an object is similar to another. There are numerous formal treatments that provide theoretical basis and empirical similarity measurement. Four approach of similarity have been proposed:

4.1. Geometric models

The first cognitive approach to comparison is known as the mental distance approach [19]: two entities (or two concepts) are represented as point in a dimensional space. The more the points are near to each other, the more entities are similar.

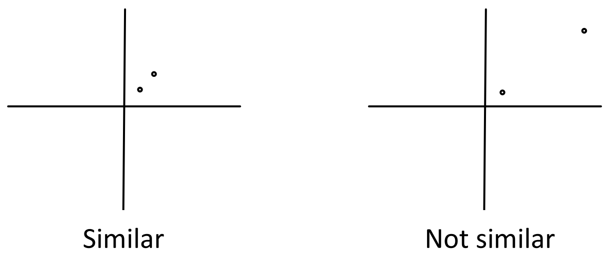

Figure 2: Schema of mental distance approach of similarity

This approach offers many opportunities for deriving space from data: Multidimensional Scaling [19] and more recently Latent Semantic Analysis [28]. Multidimensional scaling (MDS) consists in positioning two entities in a dimensional space where similar entities will be close to each other. For example, if we ask people to rate similarity from 1 to 10 of Russia, Cuba, Jamaica, we find [20]:

similarity(Russia,Cuba)=7,
similarity(Cuba, Jamaica)=8,
similarity(Russia, Jamaica)=1.

An application of MDS is to determine the underlying dimensions of the chosen space: in the example, the dimension to compare Russia and Cuba might be the political affiliation and the dimension to compare Cuba and Jamaica is probably the climate or geographical position. This model has been applied with success to express cognitive structures in classical chess position [21] and in flight scenario [22]. This approach uses a space to represent entities, and a distance to compare entities. This approach has some limitations: distances rules are not always fulfilled: minimality is not fulfilled if identical objects have not the same similarity; symmetry is not fulfilled if an object is judged similar to another more than the reverse – North Korea is judged similar to China more than China to North Korea.

4.2. Featural models

The second cognitive approach on comparison is the features based approach [23].

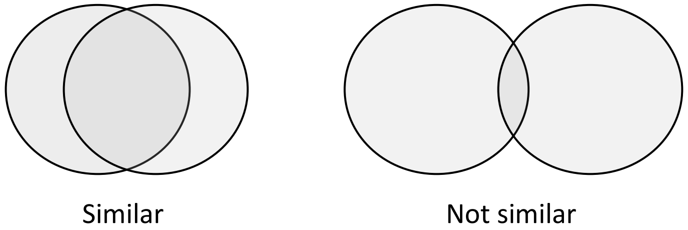

Figure 3: Schema of feature based approach of similarity

This approach has been developed to respond to lift some limits of geometric approach: the difficulty of representing two objects defined by a lot of dimensions, the strict limitation of the number of neighbors an object can have. In this approach, entities are considered as sets (or lists) of features that describe the properties of entities. A similarity comparison involves comparing the feature sets that represent the compared entities. The shared features are commonalities and the distinct features are differences. The similarity between two features depends on the number of commonalities and differences between them. For example, to compare an automobile and a motorcycle, the similarity can be defined as a ratio of what is similar to what is different. A motorcycle and a car are vehicles, they have wheels, horns, motors in common ; but motorcycle have two wheels whereas a car have four, a car have a windscreen, the motorcycle does not. This approach as lead to the contrast model where entities are represented as a collection of features and similarity is computed by: $S(A,B) = \theta f(A \cap B) - \alpha f(A - B) - \beta f(B - A)$

The similarity of A to B is expressed as a linear combination of the measure of the common and distinctive features. $A \cap B$ represents the features that items A and B have in common. (A-B) represents the features of A that are not in B, (B-A) represents the features of B that are not in A. θ, α, β are weights for the common and distinctive components.

The fundamentals premises of this approach, is to consider entities as a set of constituent features, with some of them increasing similarity and some others decreasing similarity between the compared entities.

4.3. Alignment-based Models

Geometric and features based approaches are not adapted to compare complex entities [24]. A third approach on comparison has emerged from a research on analogy: the structural approach [25].

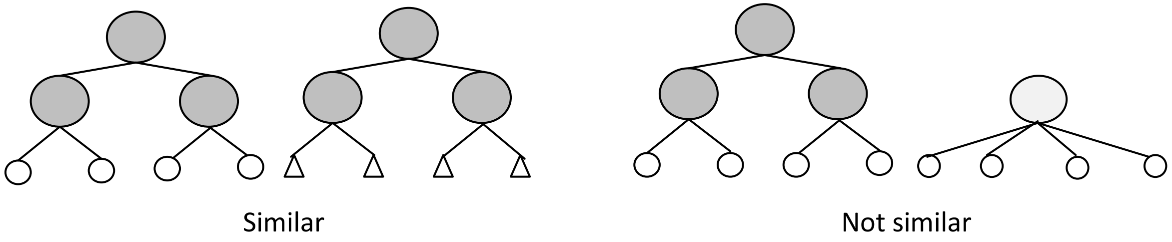

Figure 4: Schema of structural approach of similarity

This approach is used to compare two structured entities: two structured entities are similar if one element of the first entity corresponds to one element of the second entity with a connectivity constraint: the relation between the element in the first entity and other elements of this entity have to correspond to the relation between the element of the second entity and other elements of this entity: element and relations between elements have to correspond. This approach allows the comparison various kinds of entities such as stories, sounds and pictures. The fundamental aspect is to take into account the constituent part of entities (attributes) and the relations between them.

4.4. Transformational models

A fourth and original approach has been proposed: the transformational approach [26]. In this approach, two entities are similar if there is a few number of transformations needed to transform the first entity into the second one. For example: to transform “2,4,6,8” into “1,2,3,4”, only one operation is needed (divide by two); transforming “3,5,7,9” into “1,2,3,4” calls for two operations (subtract 1 and divide by two) so “2,4,6,8” is more similar to “1,2,3,4” than “3,5,7,9”. The complexity of a transformation is determined in accord with Kolmogorov complexity theory according to which the complexity of a representation is the length of the shortest computer program that can generate that representation [20]. This approach is different from the others because it permits to evaluate similarity independently from representation.

[20] is a good survey for further readings on similarity. In the same way [27] is a good survey of works on analogical reasoning and structural approach.

5. Our Taxonomy

As we said in the IRS pre-requisite part, to situate retrieval model in taxonomy is not always easy. We wondered how to modify an existent taxonomy to introduce the structural aspect in IRS.

Modifying [5] is not possible because the term "structural" is already used for structured documents, the use of "relational structure" might be a solution, but adding a new attribute like "exploiting neighborhood" or "exploiting structure" would be a better solution and permits to avoid the term conflict. Modifying [6] by adding an inter-document dependencies is not satisfying because some models like [15] may use inter-term dependencies as well as inter-document dependencies and have to be situated in two different places in the taxonomy. We want to propose a taxonomy which does a partition of IRS models according to their choice of similarity. Similarity model in cognitive science are classified by criterions used to decide if and how an object is similar to another. There are four different philosophies:

- Two entities are similar if they can be compared in term of distance in a space that permits it –spatial similarity.
- Two entities are similar if they share more commonalties than dissimilarities –attribute based similarity.
- Two entities are similar if they share similar attributes and relations between attributes are similar –structural similarity.
- Two entities are similar if few operations are needed to transform the first into the second –transformational similarity.

Spatial similarity and attribute based similarity are similar because they use attributes of compared entities to evaluate similarity. Structural similarity introduces the use of relation between attributes of compared entities to evaluate similarity. Transformational approach is different from previous approaches, because it is independent from representation and constitutes a generical approach: it is necessary to define the transformation and also the complexity of it.

We found that an analogy can be drawn between some IRS model and some similarity model. We choose to base our taxonomy on the three different approach of similarity – spatial, attribute-based and structural.

We choose to classify models that use a distance as similarity measure in the "distance" category. Vector space model and LSI model belong to this category. We choose to classify models using a set-theoric approach in the feature based category because they both share the fact of considering entities as sets of similar and distinctive features. Boolean models belong to this category. Concerning probabilistic and language based approach we interrogates ourselves where to place such IR models. We choose to consider the use of indexing terms as basis in the similarity calculation so we have classified them in the feature based category. We choose to place IR models using document collection structure to compare two documents in the structural category.

Analogy between IR and cognitive science gives taxonomy of IR models with a cognitive point of view:

Table I: Taxonomy of retrieval models with a cognitive point of view

Distance	**Features Based**	**Structural**
Vector space LSI	Boolean Fuzzy sets Probabilistic Neural networks Inferential networks Language models	The language model used in [14] The Graph model used in [8]

Taxonomy presented in table 1 builds bridges between two universes, and helps to find out how a concept in one discipline can be translated into a concept in the other one. Furthermore, this taxonomy establishes class of IR models and then, we can study how to translate concept between different models belonging to the same class. For example, the language model used in [14] which may be classified in the language model family which belongs to the probabilistic model family is classified here as a structural approach. By the same, the graph model used in [8] which may be considered as an extension of vector model because a document-term graph is equivalent to a document-term matrix. In our taxonomy this model is typically structural because it uses the features of compared document as well as both relation between indexing terms, and relation between documents during the similarity calculation.

We note that no IR model (as far as we know) have a similar approach with transformational models. To define a transformational approach some concept must be defined: transformation has a direction: a first entity is transformed into a second, or a second is transformed into the first. A transformation has a complexity that need to be defined.

Some may say that IR models can be seen with a transformational point of view:

Vector space approach may be considered as a transformational approach too: the relevance of a document for a given query is proportional to the distance between them that can be seen as the smallest geometrical translation needed to transform a document into the query

Language model can be seen as a transformational approach: a document is similar to a query if the document generates the query by series of translation. In fact, this model seems to exploit transformation notion, but the similarity measure is not determined by the features of the transformation but by a statistical count of transformations. In addition if we choose to place language model approach in the transformational category, probabilistic model cannot be placed in it because of a difference in the transformation direction.

KL-divergence approach might be seen as a transformational approach: the approach that transforms a language model into another with the smallest amount of difference. In the similarity calculation, the indexing terms distribution is compared. We choose to retain the fact that the indexing terms are used, so we classify all probabilistic models in feature based in order to oppose them to structural models.

. This absence of a proper definition of the transformational approach in IR and the complexity of this approach leads us to ask ourselves about the opportunity of an IR model using transformational approach.

An IR model based on transformation has still to be defined. This call for defining what transformation can be, and what complexity it may produce.

6. Conclusion

Working on a graph based model to exploits relation between documents and terms, relations between terms and relations between documents, we found ourselves wondering how to map our approach into an existent IR taxonomy. We studied some taxonomies, and tried to situate our approach into them. We encountered some problems, especially the one related to the fact that relational aspects are not addressed. We tacked comparison in a more general field: cognitive science where researchers have studied the criteria used to define similarity between entities.

We found an analogy between automatic comparison in IRS and general comparison in cognitive science. This analogy led us to propose a new taxonomy which classify IR model following the kind of similarity used to decide if an object (a document) is similar to another (a query) or not. In our taxonomy, IR approaches that belong to each class are associated to be the same cognitive approach of performing comparison: for example, set theoretic approach and probabilistic approach are related with regard to similarity. This fact might be considered as a limitation, and calls us to wonder how to integrate a way to differentiate theses models into our classification. A possible solution is to integrate a second dimension in our taxonomy like mathematical basis of the model.

The main advantage of this taxonomy is to highlight what we call the structural aspect: the use of features of compared entities (as in the features based approach), the use of relation between features (indexing terms) and the relations between entities (compared documents). We found two structural IR models: one issue from the vector space approach and the other from the language model approach. We think that the structural aspect is becoming more and more important, and structural notion may appear in existent approach.

Nevertheless, analogy between IR and cognitive science is not complete: we have not found any transformational approach in IR. This absence leads us to ask ourselves about the opportunity of an IR model using transformational approach. Transformational approach may become an issue for IRS.

7. References

[1] SMITH L.C., WARNER A.J. *A taxonomy of representation in information retrieval design*, Journal of Information Science, 8, p.113–121, 1984.

[2] CROFT W.B. *Approaches to intelligent information retrieval*, Information Processing and Management, 23(4), p.249–254, 1987.

[3] NIE JIAN-YUN *Un modèle logique général pour les systèmes de recherche d'informations : application au prototype RIME*. Phd thesis University Joseph-Fourier – Grenoble, 1990.

[4] PAIJMANS H., *Explorations in the document vector model of information retrieval*, Dissertation, Tilburg University, 1999.

[5] CANFORA G., CERULO L. *A taxonomy of information retrieval models and tools*. CIT. Journal of computing and information technology, vol. 12, n°3, p. 175-194, 2004.

[6] KUROPKA D. *Modelle zur Repräsentation natürlichsprachlicher Dokumente. Ontologie-basiertes Information-Filtering und -Retrieval mit relationalen Datenbanken*. Advances in Information Systems and Management Science, Bd. *10. I*SBN 978-3-8325-0514-1, 264 pages, 2004.

[7] C.J VAN RIJBERGEN, *A non-classical Logic for Information Retrieval*. Computer Journal, vol.29, N°6, 1986.

[8] CHAMPCLAUX Y. *Un modèle de recherche d'information basé sur les graphes et les similarités structurelles pour l'amélioration du processus de recherche d'information*. Phd thesis, University of Toulouse, 2009.

[9] L. A. ZADEH *Fuzzy Sets*. Information and control, p. 338-353, 1965.

[10] KWOK, K. L. *A neural network for probabilistic information retrieval*, In *ACM SIGIR Proceedings*, p. 21-30, 1989.

[11] TURTLE, H., AND CROFT, W. B. *Inference network for document retrieval*. Research and Development on Information Retrieval - ACM-SIGIR, Brussels, p. 1-24, 1990.

[12] LANDAUER, T. K., FOLTZ, P. W., & LAHAM, D. *An introduction to latent semantic analysis*. Discourse Processes, 25, 259-284. 1998.

[13] PONTE, J. AND W. B. CROFT. 1998. *A language modeling approach to information retrieval*, In: SIGIR'98, pp. 275-281.

[14] 0REN KURLAND. *Inter-document similarities, language models, and ad hoc information retrieval*, A Dissertation presented to the Faculty of the Graduate School of Cornell University, 2006.

[15] BISSON G., HUSSAIN F.: *Chi-Sim: A New Similarity Measure for the Co-clustering Task*. ICMLA, p. 211-217, 2008.

[16] CARRIERE J. AND KAZMAN R., *Webquery : Searching and visualizing the web through connectivity*, In Proceedings of the six International World Wide Web, p 701-711, 1997.

[17] SERGEY BRIN AND LAWRENCE PAGE, *The anatomy of a large-scale hypertextual Web search engine*. Source, Computer Networks and ISDN Systems archive, vol. 30 , Issue 1-7, April 1998.

[18] J. KLEINBERG *Authorative sources in a hyperlinked environnement*, Proc. ACM-SIAM Symposium on Discrete Algorithms, 1998. Also appears as IBM Research Report RJ 10076(91892) May 1997.

[19] ROGER SHEPARD *The analysis of proximities: Multidimensional scaling with an unknown distance function*. Psychometrika, Springer, vol. 27(2), p. 125-140, June 1962.

[20] KEITH J. HOLYOAK, ROBERT G. MORRISON *The Cambridge Handbook of Thinking and Reasoning (Cambridge Handbooks in Psychology)*. Cambridge University Press. ISBN: 0521531012, 858 p., 2005.

[21] HORGAN D. D., MILLIS K., NEIMEYER R. A. *Cognitive reorganization and the development of chess expertise*. International Journal of Personal Construct Psychology, 2, p. 15-36, 1989.

[22] SCHVANEVELDT, R. *Measuring the structure of expertise*. International Journal of Man-Machine Studies, 23, p. 699-728, 1985.

[23] TVERSKY A. *Features of similarity*. In Readings in Cognitive Science from Psychological Review 84, p. 327-352, 1977.

[24] HUMMEL, J.E. *Where view-based theories break down: The role of structure in shape perception and object recognition*. In E. Dietrich and A. Markman (Eds.). Cognitive Dynamics: Conceptual Change in Humans and Machines. Hillsdale, NJ: Erlbaum, 2000.

[25] MARKMAN, A. B., & GENTNER, D. *Structural alignment during similarity comparisons*. Cognitive Psychology, 25, 431-467. 1993.

[26] HAHN, U., & CHATER, N. *Understanding Similarity: a Joint Project for Psychology*, Case-Based Reasoning and Law. Artificial Intelligence Review, 12, p. 393-427, 1998.

[27] FRENCH, R. M. *The Computational Modeling of Analogy -Making*. Trends in Cognitive Sciences, *6*(5), p. 200-205, 2002.

[28] G. FURNAS, S. DEERWESTER, S. DUMAIS, T. LANDAUER, R. HARSHMAN, L. STREETER, AND K. LOCHBAUM. *Information retrieval using a singular value decomposition model of* latent semantic structure. In Proceedings of ACM SIGIR 88, p. 465–480, 1988.

[29] LANCASTER, F.W. AND FAYEN, E.G. *Information Retrieval On-Line*. Melville Publishing Co., Los Angeles, California, 1973.

[30] J. BALINSKI AND C. DANILOWICZ. *Re-ranking method based on inter-document distances*. Information Processing and Management, 41(4):759–775, 2005.

[31] O. KURLAND AND L. LEE. *PageRank without hyperlinks: Structural re-ranking using links induced by language models*. In Proceedings of SIGIR, pages 306–313, 2005.

[32] INGWERSEN P., *Intermediary function in Information retrieval interaction.* Copenhagen Buisness School. Faculty of Economics, Compenhagen : Samfundsliterrature. Doctoral dissertation, 1991.

[33] SPARCK JONES, K., Walker, S. and Robertson, S.E. (1998) *A probabilistic model of information retrieval: development and status.* TR 446, Computer Laboratory, University of Cambridge

[34] SALTON, G., FOX, E. A., & WU, H. (1983). *Extended Boolean information retrieval.* Communications of the ACM, 26, 1022-1036.

[35] G. SALTON , A. WONG , C. S. YANG, *A vector space model for automatic indexing*, Communications of the ACM, v.18 n.11, p.613-620, Nov. 1975.