Context – Sensitive Query Expansion Based on Fuzzy Clustering of Index Terms

Giorgos Akrivas, Manolis Wallace, Giorgos Stamou, and Stefanos Kollias

Image, Video and Multimedia Laboratory,
Department of Electrical and Computer Engineering,
National Technical University of Athens,
15773 Zografou, Greece
{gakrivas,wallace}@image.ntua.gr, {gstam, stefanos}@softlab.ntua.gr
http://www.image.ntua.gr

Abstract. Modern Information Retrieval Systems match the terms contained in a user's query with available documents through the use of an index. In this work, we propose a method for expanding the query with its associated terms, in order to increase the system recall. The proposed method is based on a novel fuzzy clustering of the index terms, using their common occurrence in documents as clustering criterion. The clusters which are relevant to the terms of the query form the query context. The terms of the clusters that belong to the context are used to expand the query. Clusters participate in the expansion according to their degree of relevance to the query. Precision of the result is thus improved. This statistical approach for query expansion is useful when no a priori semantic knowledge is available.

1 Introduction

An *Information Retrieval System (IRS)* consists of a database, containing a number of documents, an *index*, that associates each document to its related terms, and a *matching mechanism*, that maps the user's query (which consists of *terms*), to a set of contained documents. Quite often, the user's query and the index are fuzzy, meaning that the user can somehow supply the degree of importance for each term, and that the set of associated terms for each document also contains degrees of association. In this case, the returned documents are sorted, with the one that best matches the user's query returned first [1].

It is possible that a query does not match a given index entry, although the document that corresponds to the entry is relevant to the query. For example, a synonym of a term found in a document may be used in the query. This problem, which is known as *word mismatch* [2], has been dealt with in the past with various methods [3], many of which have an iterative [4],[5],[6] or an interactive [7] manner. Such approaches rely on relevance feedback.

Still, the dominating approach to the problem of word mismatch is the use of a thesaurus containing, for each term, the set of its related ones. The process of enlarging the user's query with the associated terms is called *query expansion*;

it is based on the associative relation A of the thesaurus, which relates terms based on their probability to co-exist in a document [8],[9]. In such approaches, as is for example the LCA method described in [10], the thesauri are generally created based on term co-occurrences in documents. Unfortunately, as has been shown on [11], the simple use of a general thesaurus for query expansion provides limited improvement. This can be justified as follows: query expansion by using a thesaurus results in the selection of numerous documents that are related to the query terms only if taken out of context.

In order to make query expansion sensitive to context, the matching mechanism must be made more intelligent. Usually, this involves a *semantic encyclopedia*, which can be used as a means to provide semantics to the user's query [12],[13],[14]. The semantics are used to extract the query context, which is subsequently used to direct query expansion towards terms that are related to the context. Such a method has been proposed by the authors in [16].

However, a semantic encyclopedia is not always available, because it requires tedious and time-consuming labor by a human expert. Moreover, the scope of the encyclopedia might be unrelated to some of the terms. In this work, we propose a method that extracts the query context using the index of the IRS, instead of a knowledge base. The method is as follows: A fuzzy clustering of the index terms, based on statistical measurements on the index, extracts a set of possible query contexts. These clusters are compared with the query and the context of the query is expressed on terms of clusters. The extracted context is subsequently used to direct query expansion towards the clusters that are related to it.

The paper is organized as follows: In section 2, we provide the mathematical framework. In section 3 we provide the algorithm for fuzzy clustering of index terms. In section 4.1 we use the clusters of terms to detect the context of the query. In sections 4.2 and 4.3, we use context to expand the user's query in a meaningful way. Finally, sections 5 and 6 provide a simulation example and our final conclusions, respectively.

2 Mathematical Framework

Before continuing, we provide the reader with the mathematical framework on fuzzy sets and relations.

Let $S = \{s_1, s_2, \ldots, s_n\}$, where $n = |S|$, denote the set of indexed terms.

A *fuzzy set* q on S is a function $q : S \to [0, 1]$. Its *scalar cardinality* is defined as $|q| = \sum\limits_{s \in S} q(s)$. From now on, for the fuzzy set q, we will use the sum notation $q = \sum\limits_{i \in \mathbb{N}_n} s_i/q_i$, where $q_i = q(s_i)$.

The subset relation, for fuzzy sets, is defined as $q_1 \subseteq q_2 \Leftrightarrow q_1(s) \leq q_2(s), \forall s$

A *fuzzy binary relation* on $X \times Y$ is a fuzzy set on $[0, 1]$, i.e. a function $R : X \times Y \to [0, 1]$.

The *intersection* and *union* of two fuzzy relations P and Q defined on the same set S are defined as:

$[P \cap Q](x, y) = t(P(x, y), Q(x, y))$

$[P \cup Q](x, y) = u(P(x, y), Q(x, y))$
where t and u are a triangular norm and a triangular co-norm, respectively. The *standard* t-norm and t-conorm are, respectively, the min and max functions. In this work, we use the standard functions, except if noted otherwise.

The *power set* $\mathcal{P}(S)$ of S is the set of all subsets of S. $\mathcal{P}(S)$ contains 2^n members. The subset relation $\subset$ is a strict ordering relation on $\mathcal{P}(S)$.

3 Fuzzy Clustering of Indexed Terms

In this section, we propose a method for clustering of the indexed terms. Our purpose is to create clusters of terms that are highly associated with each other. The basic principle of the method is that a set of terms is considered to be a valid cluster when they are frequently encountered in the same documents. Obviously, the frequency of term co-occurrence in documents varies and consequently their association is a matter of degree; we will use the term *cluster validity* to express this degree.

According to the above, a cluster is described by the set of documents that correspond to its terms; however, a cluster term may be related to more documents than those that characterize the cluster. Terms that are contained in substantially more documents than those that characterize a cluster, must not be considered to be highly associated with it. In other words, the degree of membership a term to a cluster depends on the degree to which the documents that contain a term are limited to the cluster's characterizing documents.

Moreover, a term may appear in more than one group of documents and thus it may participate in more than one clusters. Therefore, clusters may overlap.

In the following, we will describe how the clusters can be created automatically. This process is based on an extensive indexing of a sufficient and representative set of documents of an archive; extensive indexing allows us to infer that term co-occurrence originates from their semantics. In other words, we interpret such a co-occurrence as a semantic association, which we may apply when considering the remaining documents. The degree of association depends on both cluster validity and the terms' degree of participation to the cluster.

Comparing the effort required to create the extensive index, with the one required to create a semantic encyclopedia, we can observe that indexing requires less effort because i) the thesaurus relations need not be supplied by the expert and ii) having a semantic encyclopedia does not relieve the expert from the need to index the documents.

Before continuing, we provide a few details on the index.

3.1 The Index of the Information Retrieval System

The index D is a relation that maps the set of terms S to the set of documents T. Although the index is in most cases crisp, we make the more general assumption that fuzziness is allowed. Thus, some terms are indexed to lesser degree than others. Therefore, D is a fuzzy relation $D : S \times T \to [0, 1]$.

By $D(s)$, we will denote the set of documents that contain the term s, i.e. the respective row of the index.

3.2 Structure in documents and proximity of term occurrence

Quite often, documents contain structure, in the form of subdocuments, sub-subdocuments and so on. In this case, terms that belong to the same part of the document are considered more associated than terms that belong to different parts. In this subsection, we consider how this affects the process of indexing.

Let as suppose that a document $t \in T$ contains two subdocuments, t_1, t_2. The two subdocuments are considered members of the set of documents T. Moreover, let us suppose that subdocument t_1 contains term s. Then, $D(s, t_1) = 1$ and $D(s, t_2) = 0$. The term s will be considered belonging to t to a lesser degree than 1, since it actually belongs to only one of its subdocuments. A way to estimate the importance of s in t is to use the percentage of the length of t_1 in the length of t. Thus, if t_1 is long and t_2 is short, then $D(s, t)$ will be close to 1 and vice versa.

Using the above principle, terms that belong to the same parts will have a greater degree of co-occurrence in documents than terms that just belong to the same document.

3.3 Detection of clusters

Let $\mathcal{P}(S) = \{S_1, S_2, \ldots, S_m\}$, where $m = 2^n$, be the power set of S. Each member $S_i = \{s_{ij} : j \in \mathbb{N}_{n_i}\}, n_i = |S_i|$ is a candidate cluster of terms. When we also consider degrees of association of terms to the cluster, the corresponding fuzzy cluster is:

$$S_i = \sum_{j \in \mathbb{N}_{n_i}} s_{ij}/w_{ij}$$

From now on, the term "cluster" will refer to both the crisp set and its fuzzy counterpart.

A cluster is considered valid with degree one if its members occur in the same documents; based on the assumption that the set of documents is representative, we use scalar cardinality as a measure of the degree of co-occurrence. Thus, validity shall be proportional to the number of documents that contain all terms, and inversely proportional to the number of documents that contain at least one term. Therefore, we define the validity of cluster S_i as:

$$w_i = \frac{\left| \bigcap_{s \in S_i} D(s) \right|}{\left| \bigcup_{s \in S_i} D(s) \right|}$$

In cluster S_i, term s_{ij} participates with degree one when the documents that contain it are the documents that contain all members of the cluster, i.e. when $D^*(S_i) = D(s_{ij})$. Similarly to the previous definition, we define w_{ij} as:

$$w_{ij} = \frac{\left| \bigcap_{s \in S_i} D(s) \right|}{|D(s_{ij})|}$$

Obviously, $w_{ij} \geq w_i$. Moreover, we assume that each term belongs to at least one document; therefore, the denominator in the definitions of w_{ij} and w_i will never be zero.

By defining an appropriate lower threshold c_v, and eliminating clusters S_i, for which $w_i < c_v$, we avoid non-meaningful clusters. We will use the term *validity criterion* for c_v.

However, there still may be superfluous clusters. In particular, two valid clusters $S_1 \subset S_2$ may be characterized by the same documents. In this case, S_1 is unnecessary and should be eliminated. To enable this, we define the Inclusion relation I, which is a fuzzy ordering relation on the set of clusters, i.e. $I : [\mathcal{P}(S)]^2 \to [0, 1]$. I is a subset of the subset relation defined on the clusters. Therefore, $S_i \subset S_j \implies I(S_i, S_j) = 0, \forall S_i, S_j \in \mathcal{P}(S)$. Moreover, $I(S_i, S_j)$ approaches one as w_i approaches w_j.

Based on these conditions, we define the Inclusion relation as follows:

$$S_i \supseteq S_j \implies I(S_i, S_j) = \frac{w_i}{w_j}$$

If $I(S_i, S_j)$ falls above an appropriate threshold c_m, cluster S_j will be eliminated, as redundant. We will use the term *merging criterion* for c_m.

3.4 The Clustering Algorithm

Let us now proceed with the details of the clustering algorithm. We initialize the algorithm with the singletons $\{S_{11} = \{s_1\}, S_{12} = \{s_2\}, \ldots, S_{1n_1} = \{s_{n_1}\}\}$, where $n_1 = n$.

The algorithm executes in n steps, and each step i uses as input clusters $\{S_{ij} : j \in \mathbb{N}_{n_i}\}$, with cardinality $|S_{ij}| = i$ and produces as output clusters $\{S_{i+1,j} : j \in \mathbb{N}_{n_{i+1}}\}$, with cardinality $|S_{i+1,j}| = i + 1$. Each step i executes as following:

- For each cluster $S_{ij}, j = 1, \ldots, n_i - 1$:
  - For each cluster $S_{ik}, k = j + 1, \ldots, n_i$:
    1. Compute the union $S_{ij} \cup S_{ik}$
    2. If it already exists, is invalid or has a cardinality different than $i + 1$, then it is deleted
- Delete clusters that satisfy the merging criterion

The reason we choose to limit the output of step i to clusters of cardinality $i+1$ is to ensure that the same clusters are not produced by different steps of the algorithm. Furthermore, as they will be computed again in consequent steps, this causes no loss.

By arranging clusters in a lexicographical order, we may avoid superfluous unions, as well checking for duplicate clusters and for clusters with invalid cardinality in step 2.

Both of the above contribute to the substantial reduction of the needed operations. By defining proper validity and merging criteria, and eliminating clusters that don't satisfy them, it is expected that the remaining clusters are both meaningful and non-redundant.

4 Context – Sensitive Query Expansion

As mentioned in section 1, query expansion enriches the query in order to increase the recall of the system. The presence of several terms in the query defines a context, which we use, in this section, to limit the expansion process, in order to improve the precision of the retrieval.

4.1 Detection of Context

As described in section 1, a query q is a fuzzy set defined on S. This means that any term $s_i \in S, i \in \mathbb{N}_n$ belongs to q to some degree $q_i = q(s)$. Of course, for most terms this degree is expected to be zero. Nevertheless, we assume that $q_i = 1$ for at least one term (i.e. q is normal, which means that the height is one). In this subsection, we express the context of the query in terms of the clusters that include it.

Let as first consider the case where the query is crisp. In this case $q = \{s_i\} \in \mathcal{P}(S)$. The Inclusion relation, defined in subsection 3.3, contains, for each cluster, the fuzzy set of clusters that contain it. We will use the notation $[I(S_i)](S_j) \doteq I(S_j, S_i)$ for the fuzzy set of clusters that contain cluster S_i, i.e. for the respective row of I, and the notation $I(s) \doteq I(\{s\})$ for the set of clusters that contain term s. We will use the term *context* of S_j and s for $I(S_j)$ and $I(s)$, respectively. Therefore, the context of a term contains the clusters whose terms are contained in the same documents with this term. Since we consider the terms of a cluster as having a semantic correlation, the clusters which form the context are the groups of semantically correlated terms which tend to be found in the same document with a given term, or set of terms. Thus, the context of the query can be used for query expansion.

Obviously, $S_1 \subseteq S_2 \implies I(S_1) \supseteq I(S_2)$, i.e. the presence of more terms will make the context narrower. Moreover, if $S_i = \{s_{ij}, j \in \mathbb{N}_{n_i}\}$, then $I(S_i) = \bigcap_{j \in \mathbb{N}_{n_i}} I(s_{ij})$, i.e. the overall context is the intersection of the individual contexts.

Let us now extend the above definition of context to the case of fuzzy queries. Similarly to the above definitions, we define the context $K(q)$ of the query q as a fuzzy set on the set of clusters $\mathcal{P}(S)$, i.e. $K(q) = \sum_i S_i / K(q)_i$.

First, we show that a direct extension of the above definition in the fuzzy case, for example $K^*(q) = \bigcap_i q_i K(s_i)$, is not meaningful [16]. A low degree of importance q_i for the term s_i implies that the meaning of s_i is relatively insignificant for the query. On the other hand, it is implied by the above definition of K^* that a low value of q_i will narrow the context more than a high one; this is the opposite effect than what is desired.

In the following, we define the context $K(q)$ of the query q as the fuzzy intersection of the individual weighted contexts of each term :

$$K(q) = \bigcap_i K_q(s_i)$$

In order to achieve the desired effect, i.e. having a query term with low weight narrowing the context less than one with a hight weight, the following conditions must be satisfied, for the weighted context $K_q(s_i) = \sum_j S_j / K_q(s_i)_j$ of term s_i:

- if $q_i = 0$, then $K_q(s_i) = K(q \cap (S \backslash \{s_i\}))$ (no narrowing of context)
- if $q_i = 1$, then $K_q(s_i) = I(s_i) \cap K(q \cap (S \backslash \{s_i\}))$
- $K_q(s_i)_j$ decreases monotonically with q_i

Our approach is linear:

$$K_q(s_i)_j = 1 - q_i(1 - I(s_i)_j)$$

We will use this definition of context in subsection 4.3 in order to perform a context - aware query expansion. When the context contains high degrees for the clusters, the context will be important for query expansion. As a measure of this importance, we will use the height of the context $l = h(K(q))$; we will use the term *intensity* for l.

4.2 Handling of Terms in Query Expansion

In section 1, we explain that the search engine uses the query q and the document index D, which is a fuzzy relation between the set of terms S and the set of documents T, to produce the result r; r is a fuzzy set on T. When the query is comprised of a single term s, then the result is simply the respective row of D, i.e. $r(q) = D(s)$. When the query contains more than one terms, then the result is the set of documents that contain all terms, i.e. $r(q) = D(s_1) \cap D(s_2) \cap D(s_3)$.

In query expansion, we replace each term s with a set of terms $X(s)$; we will refer to this set as the *expanded term*. During evaluation of the query, we treat $X(s)$ considering a union operation, i.e. documents that match any term contained in $X(s)$ are selected. Therefore, in order to preserve the intersection operation among the original query terms, we need to expand each term separately.

4.3 Term Expansion

Using the above principle, we define, in this subsection, the expansion $X(s_i)$ of a single query term s_i as a fuzzy set on s, i.e. $X(s_i) = \sum_j s_j / x_{ij}$. We define it as a function of the query q, the context $K(q)$ of the query, and the set of valid clusters. The weight x_{ij} denotes the degree of significance of the term s_j in $X(s_i)$.

In a context – insensitive query expansion, i.e. when the intensity l of the query context is zero, the weight q_{ijk} must increase monotonically with respect to the following quantities, with respect to a cluster S_k:

- the weight q_i of term s_i in the query.
- the weight w_{ki} of term s_i in S_k
- the weight w_{kj} of term s_j in S_k

Based on the above, we define:

$$q_{ijk} \doteq q_i \cdot \min(w_{ki}, w_{kj})$$

For each cluster Sk, q_{ijk} denotes the minimum degree, to which s_j must participate in $X(s_i)$. Therefore, the degree with respect to all clusters, in a context – insensitive expansion, is:

$$q_{ij} = \max_{k \in \mathbb{N}_m} \{q_{ijk}\}$$

In a context – sensitive query expansion, the weight x_{ijk} increases monotonically, with respect to the degree to which the context of S_k is related to the context of the query. We will use the quantity

$$l_k \doteq \frac{h(K(q) \cap I(S_k))}{l}$$

as a measure of this relevance.

The following conditions must be satisfied by x_{ijk}:

- x_{ijk} increases monotonically with respect to q_{ijk}
- $l_k \to 1 \implies x_{ijk} \to q_{ijk}$
- $l = 0 \implies x_{ijk} \to q_{ijk}$
- $l \to 1 \implies x_{ijk} \to q_{ijk} l_k$

Following a linear approach, we obtain:

$$x_{ijk} \doteq (1 - l^{\alpha}(1 - l_k))q_{ijk}$$

α is a positive parameter which controls the importance of context in query expansion. As α approaches zero, context is considered to the greatest extent. As α approaches infinity, the query expansion process becomes insensitive to context.

We can observe that the expanded query is a superset of the original one, regardless of the context intensity.

	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19
A	1	1	1	1	1	1	1	1	1	1	1	1	1				1	1	1
B				1	1	1					1	1							
C	1	1	1				1	1					1						
D									1	1							1	1	1
E				1	1	1						1							
F							1												
G								1											
H	1	1											1						
I	1	1	1	1	1	1								1	1	1			
J	1	1	1												1	1			
K	1			1	1	1								1					
L	1	1	1												1	1			

Table 1. The index. Rows correspond to terms and columns to documents. Blanks correspond to zeros.

5 Simulation Example

In order to simulate the proposed methods, we provide an example of an index in Table 1. The documents of the index regard engines and airplanes. The index terms are: A="engine", B="rocket", C="internal combustion", D="external combustion", E="turbine", F="two–stroke", G="four–stroke", H="Diesel", I= "airplane", J="propeller", K="jet", L="propeller airplane".

The clusters of terms that our algorithm detects, excluding the singletons, are: AB,AC, AD, AI, AK, BE, CF, CG, CH, HK, IK, JL, ABE, ABK, ACH, ACI, AIK, BEI, BEK, BIK, CHK, CJL, HJL, HKL, IJL,ABEI, ABIK, AJKL, CHJL, CIJL, HIJL, IJKL, ACIJL, CHIJL, HIJKL, ACHIJL. The thresholds were $c_v = 0.1$ and $c_m = 0.9$.

Let the query be $q = A/1 + I/1$. The context–insensitive expansion of q is: $x_A = A/1 + B/0.25 + C/0.38 + D/0.31 + E/0.25 + H/0.19 + I/0.38 + J/0.16 + K/0.24 + L/0.16$ and $x_I = A/0.38 + B/0.27 + C/0.25 + E/0.27 + H/0.20 + I/1 + J/0.56 + K/0.56 + L/0.56$

The intensity of the query context is $l = 0.38$. The context–sensitive expansion of q, with $\alpha = \frac{1}{3}$ is: $x_A = A/1 + B/0.20 + C/0.27 + D/0.10 + E/0.16 + H/0.10 + I/0.38 + J/0.12 + K/0.19 + L/0.12$ and $x_I = A/0.38 + B/0.22 + C/0.16 + E/0.22 + H/0.12 + I/1 + J/0.36 + K/0.43 + L/0.36$.

It can be observed that term D="external combustion", which is related to engines but not in the context of airplane engines, is diminished in the context–sensitive expansion. This is derived from the index, which does not contain documents with both airplanes and external combustion engines.

6 Conclusions and Future Work

In this work, we consider term co-occurrence in documents in order to form groups of correlated terms. This clustering does not need as input the number of the clusters, neither does it produce a crisp hierarchy. We express the context of the query in terms of clusters and use it to expand the query in a context–sensitive manner. This statistical approach is useful when no knowledge about the terms is available.

In the proposed method, mainly linear approaches are applied, for the sake of simplicity. We believe that more general, non – linear approaches might be interesting to investigate. Moreover, the non-exponential computational complexity of the clustering algorithm must be confirmed. Finally, the result of the statistic query expansion can also be combined with the result of the semantic query expansion in an efficient manner.

References

1. Kraft D.H., Bordogna G. and Passi G., Fuzzy Set Techniques in Information Retrieval, in James C. Berdek, Didier Dubois and Henri Prade (Eds.) *Fuzzy Sets in Approximate Reasoning and Information Systems*, (Boston: Kluwer Academic Publishers, 2000).
2. G. Salton, M.J. McGill, Introduction to Modern Information Retrieval, McGraw-Hill, New York, 1983.
3. Efthimiadis, E. N. Query expansion. In M. E. Williams (Ed.), *Annual review of information science and technology*, (vol. 31, pp. 121-187). Medford NJ: Information Today Inc. 1996
4. Harman, D. K. (1992). Relevance feedback revisited. In N. J. Belkin, P. Ingwersen and A. Mark Pejtersen (Eds.), SIGIR 92, *Proceedings of the 15th annual international ACM SIGIR Conference on research and development in information retrieval* (pp. 1-10). New York: ACM.
5. Salton, G. and Buckley, C. (1990). Improving retrieval performance by relevance feedback. *Journal of the American Society for Information Science*, 41, 288-297.
6. W. Bruce Croft, Jinxi Xu, Corpus-specific stemming using word form co-occurence, in: *Proceedings of the Fourth Annual Symposium*, 1994.
7. N. J. Belkin, C. Cool, D. Kelly, S. -J. Lin, S. Y. Park, J. Perez-Carballo and C. Sikora, Iterative exploration, design and evaluation of support for query reformulation in interactive information retrieval, *Information Processing & Management*, Volume 37, Issue 3, May 2001, Pages 403-434
8. Miyamoto S., *Fuzzy sets in information retrieval and cluster analysis*, (Dordrecht/Boston/London: Kluwer Academic publishers, 1990)
9. Wen-Syan Li and Divyakant Agrawal, Supporting web query expansion efficiently using multi-granularity indexing and query processing, *Data & Knowledge Engineering*, Volume 35, Issue 3, December 2000, Pages 239-257
10. Xu, J. and Croft, W. B. (1996). Query expansion using local and global document analysis. In H.-P. Frei, D. Harman, P. Schauble, & R. Wilkinson (Eds.), SIGIR '96, *Proceedings of the 19th annual international ACM SIGIR Conference on research and development in information retrieval* (pp. 4-11). New York: ACM
11. E.M. Voorhees, Query Expansion Using Lexical-Semantic Relations, in: *Proceedings of the 17th Annual International ACM SIGIR Conference*, Dublin, Ireland, 1994.
12. Kraft D.H., Petry F.E., Fuzzy information systems: managing uncertainty in databases and information retrieval systems, *Fuzzy Sets and Systems*, 90 (1997) 183-191, Elsevier.

13. Akrivas G., Stamou G., Fuzzy Semantic Association of Audiovisual Document Descriptions, *Proc. of Int. Workshop on Very Low Bitrate Video Coding (VLBV)*, Athens, Greece, Oct. 2001
14. Akrivas G., Stamou G. and Kollias S., Semantic Association of Multimedia Document Descriptions through Fuzzy Relational Algebra and Fuzzy Reasoning (submitted)
15. Klir G. and Bo Yuan, *Fuzzy Sets and Fuzzy Logic, Theory and Applications*, New Jersey, Prentice Hall, 1995
16. Akrivas G., Wallace M., Stamou G. and Kollias S., Context – Sensitive Semantic Query Expansion, *IEEE International Conference on Artificial Intelligence Systems AIS-02* (to appear)
17. ISO/IEC JTC 1/SC 29 M4242, *Text of 15938-5 FDIS Information Technology – Multimedia Content Description Interface – Part 5 Multimedia Description Schemes*, October 2001.