FUZZY SEMANTIC ASSOCIATION OF MULTIMEDIA DOCUMENT DESCRIPTIONS

G. Akrivas, G. Stamou

National Technical University of Athens, Department of Electrical and Computer Engineering, Image, Video and Multimedia Laboratory (IVML/ECE/NTUA)

ABSTRACT

According to the emerging MPEG-7 standard, the semantic description of multimedia documents is expressed in terms of semantic entities such as objects, events, concepts, and fuzzy relationships among them. The above semantic entities can be used as index terms, in order to support the semantic search process. In this paper, we propose a method that associates the semantic entities with the aid of a fuzzy thesaurus. The proposed method is based on fuzzy relations and fuzzy reasoning.

1. INTRODUCTION

Lately, there is a growing interest in the representation, storage and retrieval of multimedia documents containing textual as well as audiovisual information. The need for *content – based retrieval* has resulted in a more sophisticated description of multimedia documents. Multimedia *content* is composed of a set of *features*, which, unlike audiovisual data themselves, are recognized as meaningful by humans. For a particular multimedia description, a set of *descriptors* is defined. The queries and the consequent search refer to specific features of the content. Documents whose features match the user query, in terms of the corresponding descriptors, are retrieved [6].

MPEG-7 becomes a standard for the description of multimedia content, designed to address the above requirements [1, 4 pp 688-695]. A large number of MPEG-7 – compliant multimedia descriptions is expected to appear in the following years. The standard has defined a number of audiovisual descriptors, in a language based on XML Schema [5], and has organized them using a special kind of complex type, the Description Scheme [4, pp. 748-759].

The DSs of MPEG-7 distinguish between structural (syntactic) and conceptual (semantic) aspects of the description [1, ch. 11-12]. Structural aspects express a low-level and machine-oriented kind of description, since they describe information such as signal segments and their properties. On the other hand, conceptual aspects express a high-level and human oriented kind of description, since they deal with semantic entities, such as objects, events and concepts. These differences result in different approaches both for the creation of the description (automatic extraction vs. annotation, possibly machine-aided) and for the query form (query by example vs. traditional term-based queries).

Usage of conceptual description has advantages over structural description, because of its proximity to human understanding of multimedia information. Although automatic extraction of semantic entities from audiovisual data is not yet suitable for applications, research in this area provides improved techniques that extract some specific syntactic features, which certain methods can correlate to semantic ones. Progress in this area, and, consequently, a large amount of semantic multimedia descriptions, is expected to be available in the following years.

In the process of semantic entities extraction, the human expert (annotator) plays an important role [7]. However, exploitation of the human analyzing capabilities has disadvantages, partly because it is tedious and time consuming and partly because it suffers of subjectivity and partiality, since different people will often consider different aspects of the content as important, and an exhaustive annotation is considered unrealistic.

Another difficulty arises from the multitude of relations defined in MPEG-7. While these enable the annotator to provide a sufficiently accurate description, they present a problem to the user, who usually desires simple forms of queries.

Textual information retrieval systems have faced the above problems with the aid of fuzzy sets theory, introducing the concept of the fuzzy thesaurus [1, 11, 12]. A fuzzy thesaurus consists of associations among pairs of semantic entities. By using the thesaurus, the user query can be expanded to contain all the associated semantic entities. The expanded query is expected to retrieve more relevant documents, because of the higher probability that the annotator has included one of the associated entities in the description.

We present, in this paper, a principle for automating the construction of a fuzzy thesaurus out of a set of semantic descriptions. This method employs a fuzzy relational system to enrich an existing conceptual description (thus reducing the work required by the annotator), and to construct a fuzzy thesaurus based on it. The system operates by applying a set of fuzzy relation operations to the semantic description. These result in the semantic description being enriched and in the associations that consist the fuzzy thesaurus being constructed. A uniform appliance of rules enforces the consistency of the data, partly correcting for the annotator's partiality and subjectivity.

The paper is organized as follows. In section 2, we present the MPEG-7 semantic entities and relations. In section 3 we elaborate the role of the fuzzy thesaurus in the process of semantic search. In section 4 we present the proposed algorithms for enriching a semantic description and constructing the fuzzy thesaurus from it. Finally, in section 5 we give an example of the operation of the system.

2. MPEG-7 SEMANTIC ENTITIES AND RELATIONS

In accordance to the MPEG-7 standard, we shall refer to the reality, in which a description makes sense, as a *narrative world*, and denote it with the symbol $W = S \cup R$. The set of all narrative worlds shall be called *universe* and denoted by $U = \{w_1, w_2, \cdots, w_{|U|}\}$.

The description tools that compose a narrative world are the set of *semantic entities* $S = \{s_1, s_2, \cdots, s_{|U|}\}$ and the set of *semantic relations* $R = \{r_1, r_2, ..., r_{|R|}\}$.

The set of semantic entities is further partitioned into the sets of objects O, events E, concepts C, places P, times T, and states A. Objects and events correspond, somewhat loosely, to nouns and verbs of natural language, respectively. It is important that there exist in the same set both specific semantic entities (*instances*, e.g. "George", "Spanish soccer team") and classes of objects, (*formal abstractions*, e.g. "human", "soccer team"). The formal abstractions are said to have an *abstraction level* of 1. Finally, the set C of concepts is defined as a semantic entity that cannot be described as a generalization or abstraction of a specific object, event, time, place, or state [1]. Concepts correspond to words such as "democracy" and "commerce".

The standard defines a rich set of relations among semantic entities. Moreover, the relations are fuzzy, as introduced in paragraph 5. The defined relations fall into the categories shown in Table 1.

Table 1: Categories of MPEG-7 semantic relations

Sets	Example
$S \times S$	ExampleOf, e.g. "George" is example of "human"
$S \times U$	Depicts, e.g. "Gone With the Wind" depicts "American Civil War"
$U \times U$	InterpretationOf, e.g. "War Crimes" and "Strategy" are interpretations of "Hiroshima Bomb"
$O \times O$	MemberOf, e.g. "player" is member of "team"
$O \times E$	AgentOf, e.g. "player" is agent of "kick"
$E \times E$	ResultOf, e.g. "goal" is result of "kick"
$C \times S$	PropertyOf, e.g. "democracy" is property of "United Kingdom"
$P \times S$	LocationOf, e.g. "Paris" is the location of "game"
$P \times P$	North, e.g. "London" is North to "Paris"
$T \times S$	TimeOf, e.g. "25th Olympiad" is time of "Dimas World Record"
$T \times T$	Before, e.g. "Queen Victoria's reign" is before "WWI"

The role of the annotator is to populate the sets of the semantic entities and relations.

We present, for the purpose of exemplifying the principles already mentioned, a short semantic description in tables 2 and 3. The example, which originates in [1], describes an abstract narrative world concerning soccer.

Table 2: Semantic entities

soccerGame-ev	period-ev
soccerGame-ob	episode-ev
sports-ev	goal-ev
AmerFootballGame-ev	kick-ev
baseBallGame-ev	team-ob
period1-ev	player-ob
period2-ev	ball-ob
referee-ob	arbiter-ev
athlete-ob	goalkeeper-ob
football-ev	

Table 3: Semantic relations

Relation	Source	Target	degree
instrumentOf	ball-ob	soccerGame-ev	1
AgentOf	player-ob	kick-ev	1
PatientOf	ball-ob	kick-ev	1
resultOf	goal-ev	kick-ev	0.8
MemberOf	player-ob	team-ob	
SpecializationOf	soccerGame-ev	sports-ev	1
IdentifiedWith	soccerGame-ev	soccerGame-ob	1
AgentOf	referee-ob	arbiter-ev	1
patientOf	soccerGame-ob	arbiter-ev	1
similarTo	AmFootballGame-ev	soccerGame-ev	0.8

similarTo	AmFootball-Game-ev	baseball-Game-ev	0.8
similarTo	baseBallGame-ev	soccer-Game-ev	0.4
partOf	period-ev	soccer-Game-ev	1
specializationOf	period1-ev	period-ev	1
specialization	period2-ev	period-ev	1
partOf	episode-ev	period-ev	1
specializationOf	goal-ev	episode-ev	1
specializationOf	kick-ev	episode-ev	1
specializationOf	player-ob	athlete-ob	1
specializationOf	goalkeeper-ob	player-ob	1
equivalentTo	football-ev	soccer-Game-ev	0.8
equivalentTo	football-ev	Amer-Football-ev	0.8

3. THE SEMANTIC ENCYCLOPAEDIA AND THE FUZZY THESAURUS

A multimedia database contains a set of multimedia documents, and their respective descriptions. In a semantic query, the user asks for specific semantic entities and the system returns documents whose description semantically "contains" them. The user query can be composed of either direct references to the semantic entities and relations, or of words, or, more generally, of audiovisual data. In the latter two cases, the user's request must be transformed to a query containing semantic entities.

It is certain, however, that the annotator has not included every possible semantic entity in the description. For example, in the above example, it is mentioned that a goalkeeper is a player, and that a player is an athlete, but it is not mentioned that a goalkeeper is an athlete. Therefore, a query requesting athletes will not retrieve an event involving goalkeepers. On the other hand, inclusion of every possible semantic entity would be both non-realistic and redundant.

A solution to the problem of redundancy would be to store all the information that is common to several descriptions into a separate semantic description. In the MPEG-7 terminology, this separate description defines the narrative world, to which the specific document semantic descriptions refer. This description would contain most of the abstract entities and their relations, since they are more likely to appear in multiple descriptions, and some of the frequently encountered non-abstract entities. This set of common semantic entities and relations is considered to contain knowledge concerning not only multimedia descriptions already found in the system, but also new descriptions, still to come. This is a semantic description, in the MPEG-7 sense, but it does not concern a certain multimedia document, but rather a narrative world. We will refer to this special kind of description as a *semantic encyclopaedia*. An encyclopaedia is a semantic generalization of the dictionary (the set of index terms) used by a textual information retrieval system.

The process of construction of the semantic encyclopaedia is similar to the process of construction of a semantic document description, since in both cases a number of semantic entities and relations must be defined. The requirements and difficulties related to the role of the human expert are found in this kind of description as well, but they are more crucial, in that the quality of this description greatly affects the quality of the query results. Moreover, specific semantic document descriptions are required to comply with the description of the semantic encyclopaedia, either by constructing the semantic encyclopaedia prior to the document descriptions, or by providing references to its entities a posteriori.

Even with an exhaustive description, however, it would still be time consuming, for a query processor, to expand the query terms: the multiplicity of the relations would result in multiple searches among them to find terms related to the query terms. On the other hand, the specific relation type is not need for the query expansion.

The fuzzy thesaurus has been used as a solution to the problem of the query expansion [11,12]. The fuzzy thesaurus is composed of a (relatively small) number of semantic relations that cover every possible semantic entity. Using the semantic relations, it is possible to find, for each semantic entity of the query, the set of its related entities, and expand the query with them. In our example, the query entity "Athlete" would be expanded to player.

The proposed thesaurus for multimedia databases is composed of the following relations: equivalence, inclusion, association. These will be denoted with the symbols *E, I,* A.

By using the *E* relation, which is similar to the MPEG-7 equivalentTo relation, the query is expanded to contain terms that are, in one sense or another, synonymous. Another similar relation is identifiedWith. Therefore, the *E* relation is constructed from these two relations

The *I* relation, on the other hand, implies a generalization relation. For example, if the query requests for a player, it is only natural to retrieve a description containing a goalkeeper, because it is a specialization of player. By contrast, if the query requests for goalkeeper, any description containing a player is not accepted. This assymetry is found in many of the relations defined in the standard, for example exampleOf, specializationOf, partOf, contains, refines.

Finally, the *A* relation implies that two entities are related, however loosely. This kind of relation will give terms that are less strongly related, than the other two, and it is intended to retrieve results that the other two cannot retrieve. There is no MPEG-7 relation that is associated to the *A* relation; on the contrary, most relations defined by MPEG-7 can be used, more or less, to construct it.

Since a fuzzy thesaurus is not a part of an MPEG-7 description, it must be constructed based on a set of existing description. Construction is based on the existing semantic relations, and on their implied meaning.

4. SEMANTIC DESCRIPTION ENHANCEMENT AND CONSTRUCTION OF THE FUZZY THESAURUS

In this section, we describe a system that takes as input an MPEG-7 compliant semantic description (the description of the narrative world that is contained in the semantic encyclopaedia) and produces, as output, an enhanced version of the description and the fuzzy thesaurus. Such a system will have to rely on assumptions that regard semantic relations.

Let as first provide the reader with the mathematical framework.

Let X, Y denote two crisp sets, which, for our application, we assume to be finite, i.e. $X = \{x_1, \cdots, x_m\}$ and $Y = \{y_1, \cdots, y_m\}$. A fuzzy binary relation between the two sets is defined as a function

$$R : X \times Y \to [0,1]$$

It is often convenient to represent a binary relation as a matrix:

$$\mathbf{R} = \lfloor r_{ij} \rfloor = [R(x_i, y_n)], 0 < i < m, 0 < j < n$$

In our application, the two sets contain semantic entities, and the element r_{ij} denotes the degree of association between the semantic entities $x_i \in X, y_i \in Y$. It is expected that the majority of the elements of the matrix representation of any semantic relation are zero. A matrix whose elements are mostly zero is called a sparse matrix. An efficient representation of a sparse matrix is:

$$\mathbf{R} = \{(x, y, R(x, y)) \setminus x \in X, y \in Y, R(x, y) > 0\}$$

A function $t : [0,1] \times [0,1] \to [0,1]$ is called a t-norm iff $\forall a, b, d \in [0,1]$:

$i)\quad t(a,1) = a$

$ii)\quad t(a,b) = t(b,a)$

$iii)\quad t(a, t(b,d)) = t(t(a,b), d)$

$iv)\quad b \le d \Rightarrow t(a,b) \le t(a,d)$

For a function *u,* if the first property is replaced with $u(a,0) = a$, then the function belongs to the family of the t-conorms.

Standard functions for t-norms and t-conorms are, respectively:

$$t(x, y) = \min\{x, y\}$$

$$u(x, y) = \max\{x, y\}$$

Given two fuzzy binary relations P, Q, defined on the same pair of sets $X \times Y$, their intersection, union and product with a scalar are defined, respectively:

$$[P \cap Q]_{ij} = t(p_{ij}, q_{ij})$$

$$[P \cup Q]_{ij} = u(p_{ij}, q_{ij})$$

$$[aP]_{ij} = ap_{ij}$$

where *t* and *u* denote a t-norm and a t-conorm, respectively.

The property of subset is defined as:

$$P \subseteq Q \Leftrightarrow p_{ij} \le q_{ij}, \forall i, j$$

Finaly, for two relations P, Q defined on $X \times Y$ and $Y \times Z$, respectively, their composition, with respect to a t-norm *t* (sup-t composition), is defined:

$$\left[P \overset{t}{\circ} Q\right]_{ij} = \sup_k t(p_{ik}, q_{kj})$$

For relations defined on a single set, i.e. $R : X \times X \to [0,1]$ the properties of reflexivity, symmetry and sup-t transitivity are defined.

R is called reflexive iff $R(x, x) = 1, \forall x$

R is called symmetric iff

$$R(x, y) = R(y, x), \forall x, y$$

and antisymmentric iff

$$R(x, y) > 0 \wedge R(y, x) > 0 \Rightarrow x = y, \forall x, y$$

Finally, *R* is called sup-t transitive (or simply, transitive) iff $R \overset{t}{\circ} R \subseteq R$

A transitive closure of a relation is a transitive relation that contains the original relation and has the fewest possible members. It can be proved [3] that if R is reflexive, then its transitive closure is given by the formula $R_T = R^{(n-1)}$, where n denotes the number of entities and $R^{(n)} = \underbrace{R \circ \cdots \circ R}_{n}$.

Similar operations for reflexive, anti-reflexive and symmetric properties can be defined, and they are trivial to compute:

$R(x, x) = 1, \forall x$ (reflexive)

$R(x, x) = 0, \forall x$ (antireflexive)

$R(x, y) = \max\{R(x, y), R(y, x)\}, \forall x, y$ (symmetric)

On the other hand, a similar operation for an antisymmetric relation requires one of the associations among two semantic entities to be discarded. Therefore, human intervention is needed here.

Of the various combinations of the properties defined, the properties of equivalence (reflexive, symmetric and transitive), compatibility (reflexive and symmetric) and ordering (antisymmetric and transitive) are often found in semantic relations. Examples of properties are found in table 4.

Table 4. Properties of some semantic relations.

Relation	properties
ResultOf	strict ordering
ExampleOf	antireflexive, antisymmetric
EquivalentTo	equivalence
IdentifiedWith	equivalence
Opposite	antireflexive, symmetric
Overlaps	reflexive, symmetric
dependsOn	transitive

These properties arise from the relations alone, and do not depend on the specific entities related. However, data supplied by the annotator does not always satisfy those properties. For example, the "before" relation is a total ordering, and hence is transitive. While the annotator might have stated that "American Civil War" is before "WWI" and "WWI' is before "WWII", it is not certain that he has also stated that "American Civil War" is before "WWII". A transitive closure would correct this inconsistency. Similarly, by performing the respective closures on relations that correlate entities of the same set, we enforce their consistency.

After ensuring the consistency of the semantic encyclopaedia, we proceed to construct the E relation of the thesaurus. As mentioned above, the thesaurus is constructed with two semantic relations:

$$R_e = equivalentTo \cup identifiedWith$$

Before proceeding with the construction of the other two thesaurus relations, we further enrich the semantic encyclopaedia by applying the following rule on each semantic relation R of the encyclopaedia:

$$R(a,b) \wedge R_e(b,c) \Rightarrow R(a,c)$$

This rule can be implemented as follows:

$$R \equiv R \circ R_e$$

It is easy to show that, since E is an equivalence relation, a single composition is sufficient to expand the relation.

After employing the E relation to enrich the encyclopaedia, we proceed to construct the I and A relations. As mentioned above, they are constructed from the relations of the encyclopaedia. Inclusion is constructed by all the relations that possess the property of ordering. Association is constructed by all the relations. Weights are used to control the importance of each relation. Some of the weights can be zero.

$$R_{I,A} = \bigcup_i c_i(R_i)$$

Selection of the coefficients is crucial. For example, if a particular soccer game has happened in a certain country (locationOf relation), it would be false to associate this country with soccer.

5. EXAMPLE

Let as consider the semantic description given in section 3.

$$O = \{\text{soccerGame - ob, team - ob, player - ob, ball - ob, referee - ob, athlete - ob, goalkeeper - ob}\}$$

$$E = \{\text{soccerGame - ev, period - ev, episode - ev, sports - ev, goal - ev, AmerFootball - ev, kick - ev, baseBallGame - ev, period1 - ev, period2 - ev, arbiter - ev}\}$$

By using the sparse matrix representation for the relations, we denote:

$$SimilarTo = \{(\text{AmerFootballGame - ev, soccerGame - ev}, 0.8), (\text{baseBallGame - ev, soccerGame - ev}, 0{,}4), (\text{AmerFootballGame - ev, baseBallGame - ev}, 0.7)\}$$

$$SpecializationOf = \{(\text{soccerlGame - ev, sports - ev}, 1), (\text{period1 - ev, period - ev}, 1), (\text{period2 - ev, period - ev}, 1), (\text{goal - ev, episode - ev}, 1), (\text{kick - ev, episode - ev}, 1), (\text{player - ob, athlete - ob}, 1), (\text{goalkeepr - ob, player - ob}, 1)\}$$

Since similarTo is a symmetric relation, we perform a symmetric closure on it. For example, the following member is included in the relation:

$$R(\text{soccerGame - ev, baseBallGame - ev}) \equiv = \max\{R(\text{soccerGame - ev, baseBallGame - ev}), R(\text{baseBallGame - ev, soccerGame - ev})\} = \max\{0, 0.4\} = 0.4$$

Similarly, a transitive closure of the specializationOf relation will result in the following elements being added to relation, among others:

$$R(\text{goalkeeper - ob, athlete - ob}) \equiv 1$$

The E relation of the thesaurus is constructed with the EquivalentTo and IdentifiedWith relations:

$$R_e = EquivalentTo \bigcup IdentifiedWith = \{(\text{soccerGame - ob, soccerGame - ev}, 1), (\text{football - ev, soccerGame - ev}, 0.8), (\text{football - ev, AmerFootbal - ev}, 0.8)\}$$

We notice that the term football is vague, because it associated both with soccer (European football) and with American football. Members that would make the relation reflexive and symmetric have been omitted above, for clarity.

Composition of the E relation of the thesaurus with the instrumentOf relation would result in the following member being added:

$$R(\text{ball - ob}, \text{football - ev}) = \max\{0, \cdots, \min\{1, 0.8\}\} = 0.8$$

Since the I and the A relations are constructed similarly, we will only exemplify the construction of the I relation.

$$R_I = 0.6 \mathit{InstrumentOf} \cup 0.8 \mathit{AgentOf} \cup \mathit{MemberOf} \cup \mathit{SpecializationOf}$$

6. CONCLUSIONS AND FUTURE WORK

A principle for enriching an existing semantic description and constructing a fuzzy thesaurus was demonstrated. The role of the fuzzy thesaurus is to make the user's query richer, in the same way that the relation compositions make the annotator's description richer.

Correspondence between these two processes, annotation and user query can be further enhanced. For example, query – specific rules could be designed, whose purpose would be to expand the user's query to include more semantic entities, in the hope that some of them might be found in the multimedia description database. These rules would also consist part of the knowledge base, because they would contain information on what the user means by requesting a specific semantic entity.

The implementation of such a semantic query management system is one of the goals of the EU FAETHON IST project, to which the authors participate.

7. REFERENCES

[1] ISO/IEC JTC 1/SC 29/WG 11/N3966, "Text of 15938-5 FCD Information Technology – Multimedia Content Description Interface – Part 5 Multimedia Description Schemes", March 2001, Singapore.

[2] J. F. Martínez – Trinidad and J. Ruiz – Shulcloper, "Fuzzy clustering of semantic spaces", Pattern Recognition 34, 2001, pp 783-793.

[3] George J. Klir and Bo Yuan, *Fuzzy Sets and Fuzzy Logic, Theory and Applications,* Prentice Hall, New Jersey, 1995

[4] IEEE Transactions on Circuits and Systems for Video Technology, Special Issue on MPEG-7, June 2001.

[5] W3C, XML Schema Recommendation, May 2001, http://www.w3.org/XML/Schema

[6] ISO/IEC JTC1/SC29/WG11 N4032, "Introduction to MPEG-7", March 2001, Singapure.

[7] G. Votsis, A. Drosopoulos, G. Akrivas, V. Tzouvaras and Y. Xirouhakis, "An MPEG-7 Compliant Integrated System for Video Archiving, Characterization and Retrieval", IASTED International Conference on Signal and Image Processing (SIP2000), Las Vegas, November 2000.

[8] B.-L. Yeo and B. Liu, "Rapid Scene Analysis on Compressed Video", IEEE Trans. on Circuits and Systems for Video Technology, vol. 5, pp. 533-544, Dec. 1995.

[9] N. Tsapatsoulis, Y. Avrithis and S. Kollias, "Facial Image Indexing in Multimedia Databases", Pattern Analysis and Applications, Vol. 4, Issue 2/3, pp 93-107, 2001

[10] Y. Avrithis, Y. Xirouxakis and S. Kollias, "Affine-Invariant Curve Normalization for Object Shape Representation, Classification and Retrieval", Machine Vision and Applications, Springer Verlag (to appear).

[11] Kraft, D.H., Bordogna, G. and Passi, G. Fuzzy Set Techniques in Information Retrieval in Fuzzy Sets in Approximate Reasoning and Information Systems, James C. Berdek Didier Dudas and Henri Prade (Eds.), Kluwer Academic Publishers, Boston, 2000.

[12] Sadaaki Miyamoto, "Fuzzy sets in information retrieval and cluster analysis", Kluwer Academic publishers, Dordrecht/Boston/London, 1990