

# Conceptual Object Grouping for Multimedia Document Management

Chong Deuk Lee and Taegwon Jeong

Division of Applied System Engineering, Chonbuk National Univ.

## Abstract

Increase of multimedia information in Web requires a new method to manage and service multimedia documents efficiently. This paper proposes a conceptual object grouping method by fuzzy filtering, which is automatically constituted based on increase of multimedia documents. The proposed method composes subsumption relations between conceptual objects automatically using fuzzy filtering of the document objects that are extracted from domains. Grouping of such conceptual objects is regarded as subsumption relation which is decided by  $\mu$ -cut. This paper proposes  $\mu$ -cut, FAS(Fuzzy Average Similarity) and DSR(Direct Subsumption Relation) to decide fuzzy filtering, which groups related document objects easily. This paper used about 1,000 conceptual objects in the performance test of the proposed method. The simulation result showed that the proposed method had better retrieval performance than those for OGM(Optimistic Genealogy Method) and BGM(Balanced Genealogy Method).

**Key Words :** fuzzy filtering,  $\mu$ -cut, FAS, DSR

## 1. Introduction

Several methods have been proposed to manage and service a variety of huge multimedia documents[1,2,4,5,7,8]. However, those methods have some difficulties in insertion, constitution and classification of document objects. This paper proposes a conceptual objects grouping method by  $\mu$ -cut, FAS(Fuzzy Average Similarity) and DSR(Direct Subsumption Relation), which is automatically constituted with the increment of document objects. The proposed method composes conceptual objects grouping between the document objects automatically.

The proposed method composes groups of conceptual objects which relate structurally with other conceptual objects, and the used conceptual objects are specification of those which come from query's requirement for grouping. In other words, if document objects and domains are proposed in specific query, conceptual objects are extracted from the properties of documents. Therefore, these conceptual objects are used to determine whether document objects are in the group. However, although existing methods have an advanced capability, they have a defect that requires subsumption relation between conceptual objects. The method proposed in this paper has merits of composing conceptual objects hierarchically by fuzzy filtering in addition to maintaining subsumption relation between conceptual objects systematically. The concept-based object processing scheme has been used in the areas of artificial intelligence, natural language processing, pattern recognition and information communication.

Category-based search[3] groups documents by defined category groups to cluster the documents which belong to the same category. Until now, grouping methods are mainly being used by probability method and conceptual distance which group and combine the nearest things after measuring distance between objects. This paper proposes conceptual objects

grouping method to manage and serve relevant conceptual objects, and provides multimedia informations efficiently which are required by users in the established system. Therefore, the retrieval time and the service time are reduced, the capacity of the system increases, and the usage of the system is maximized. Also, the proposed method provides more friendly approach to the system, easier way to establish relevant multimedia informations, and fast way to store document objects required by users.

This paper is organized as follows: description of grouping method in chapter 2, experimental results of grouping method in chapter 3 and conclusions in the final chapter.

## 2. Grouping determination method

The purpose of grouping is to assign related conceptual objects and instances of conceptual objects to a same group. Existing groups will be modified or a new group will be composed during this process.

A basic strategy for grouping is as follows. If properties of conceptual objects have properties of a group, they are subsumed in the group and they aren't if not. The property of each conceptual object is expressed differently. Yet, since the conceptual objects in the same group have similar properties mutually, it is possible to group related conceptual objects. Each grouping is processed as follows.

### 2.1 Grouping by $\mu$ -cut

In the process of conceptual object acquisition in document domain, if there is a subsumption relation satisfying the properties of conceptual objects, the conceptual object is assigned to the group. Generally, the grouping is determined in the process of the conceptual acquisition of document objects. MGC(Most General Concept object) among conceptual object classes determines whether a conceptual object is grouped

together and operated in connection with other conceptual objects.

Parameters which determine grouping are as the following.

Table 1. Parameters which determine grouping

Parameters	Meaning
① C(S)	object which has properties of conceptual objects.
② MGC	conceptual object which are most generalized in C(S).
③ MC(S)	conceptual object set which satisfies MGC among C(S).
④ MGC(S)	conceptual object set which carries out $\mu$ -cut with MGC among MC(S).
⑤ CPM(S)	conceptual object which satisfies $\mu$ -cut in group.
⑥ CPM(N)	conceptual object which doesn't satisfy $\mu$ -cut.

Here, a set of the satisfied conceptual objects is represented as the following.

$$MC(S) = \{CPM(S)-CPM(N)\}.$$

MGC is the conceptual object which is most generalized in C(S) and carries out  $\mu$ -cut to determine final common properties of conceptual objects.

In this paper,  $\mu$ -cut is defined as the following to determine grouping of conceptual objects:

(Definition 1)  $\mu$ -cut = { c | MC(S)  $\geq$   $\mu$  }

where, MC(S)  $\geq$   $\mu$  is a set which satisfies properties of conceptual objects in a group.

MC(S)  $\geq$   $\mu$  must satisfy following conditions.

- (1) Reflexive relation :  $\mu \approx (C_i, C_j) = 1$
- (2) Symmetric relation :  $\mu \approx (C_i, C_j) = (C_j, C_i)$
- (3) Transitive relation :  $\mu (C_i, C_k) \geq \min\{\mu$ -cut  $\approx (C_i, C_k), \mu$ -cut  $(C_j, C_k)\}$

Transitive relation is determined by  $\mu$ -cut, and  $\mu$  is a membership function with fuzzy value between 0 and 1 and determines the relationship of MC(S). Fig.1 represents a hierarchical structure by  $\mu$ -cut.

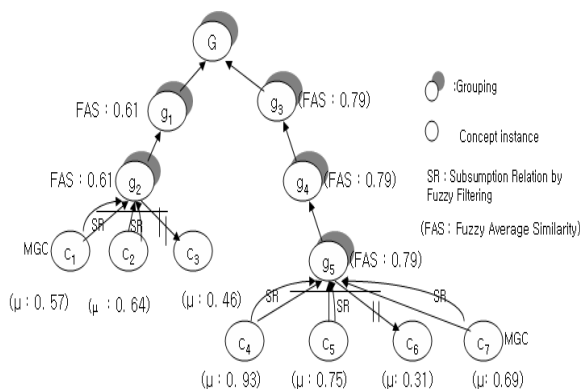


Fig. 1 Hierarchical structure by  $\mu$ -cut

Here, a set of the satisfied conceptual objects by  $\mu$ -cut is represented as a MS(C) = {CPM(S)-CPM(N)}.

"c<sub>1</sub>, c<sub>2</sub>, c<sub>4</sub>, c<sub>5</sub>, and c<sub>7</sub>" are nodes which compose the subsumption relation, but c<sub>3</sub> and c<sub>6</sub> are opposite nodes. If C(S) is in g<sub>2</sub>, g<sub>5</sub> is {c<sub>1</sub>, c<sub>2</sub>, c<sub>3</sub>, c<sub>4</sub>, c<sub>5</sub>, c<sub>6</sub>, c<sub>7</sub>} and CPM(N)=CPM{c<sub>3</sub>, c<sub>6</sub>}, then MC(S)={CPM(S)-CPM(N)} will be {c<sub>1</sub>, c<sub>2</sub>, c<sub>4</sub>, c<sub>5</sub>, c<sub>7</sub>}.

Now, if c<sub>1</sub> and c<sub>7</sub> is MGC, MGC(S)=(g<sub>2</sub>  $\vee$  g<sub>5</sub>) will be carried  $\mu$ -cut. Besides, if c<sub>3</sub> and c<sub>6</sub> are nodes which don't compose the subsumption relation, then the grouping happens because the subsumption relation is possible toward g<sub>2</sub>, g<sub>1</sub>, g<sub>5</sub>, g<sub>4</sub>, g<sub>3</sub> and G.

## 2.2 Grouping by FAS(Fuzzy Average Similarity)

This is the process to determine the subsumption relation by FAS(Fuzzy Average Similarity) when a new conceptual object is available. The grouping process must immediately determine the subsumption relation. If the FAS of a conceptual object satisfies the subsumption relation, the conceptual object is subsumed. If not, the conceptual object will be filtered.

FAS is defined as the following:

$$FAS = \sum_{G_j \text{ is a child of } G_i} (\mu\text{-cut}) = \{ C | MC(S) \geq \mu \} / Size(N)$$

Where, Size(N) is a number of conceptual objects which satisfies MC(S)  $\geq$   $\mu$ .

Based on (Definition 2), subsumption relation by FAS is achieved as the following:

Step 1. FAS processing about conceptual object of document objects in a domain.

```

procedure fuzzy_value relation()
{
  procedure fuzzy_relation(ci, cj)
  // Fuzzy-filtering processing about conceptual objects.
  procedure extend_search(ck)
  // extend searching for fuzzy filtering.
  for(i=0; i++)
  fuzzy_filtering(ci, ci+1)
  extract  $\mu$ _value()
}
    
```

Step 2. conceptual object-based grouping creation which satisfies FAS.

```

procedure extend_search()
{
  extract_  $\mu$  (membership function)()
  // create  $\mu$ -cut value for new conceptual objects.
  for(i=0; i++){
  if( $\mu$ -cut)={ C | MC(S)  $\geq$   $\mu$  }
  create FAS Group(ci)
  // create conceptual object-based grouping which satisfies FAS
  for a new conceptual object.
}
    
```

Table 2 shows parameters to determine the subsumption relation in *Step1 and Step2*.

Table 2. The parameters to determine subsumption relation in *Step1 and Step2*.

NEW(C)	conceptual objects which carries out FAS.
FAS(C)	conceptual objects which satisfies $MC(S) \geq \mu$ in NEW(C).

If FAS meets a new conceptual object that carries out the subsumption relation, a new set of conceptual object that satisfies is  $MC(S)=CPM\{(FAS(C) \vee \{C_i\})-CPM(N)\}$ . At this point, the subsumption relation is determined by the disjunction of present sets and MGC,  $MGC(S)$  among  $MC(S)$ .

If FAS meets a new conceptual object which does not satisfy the subsumption relation,  $MC(S)=\{(NEW(C)-CPM(N \vee \{C_i\}))\}$ . In determining FAS, the subsumption relation is determined by the disjunction of MGC and  $MC(S) \geq \mu$  among  $MC(S)$ .

For example, if in Fig.1 the current MGC is  $c_7$ , and a new conceptual object node of NEW(C) is  $P_1 \geq 0.69$ , then the condition is  $CPM\{MC(S) \vee NEW(C)\}=\{c_4, c_5, c_7, P_1\}$  and  $CPM(N)=\{c_6\}$ .

To determine the subsumption relation of the conceptual objects effectively, firstly, we must proceed  $MC(S) \geq \mu$  among NEW(C) which satisfy the subsumption relation and, secondly, limit the property size of a conceptual object. Therefore, if we limit the property size of a conceptual object, then the number of node to be searched decreases, and thus we can perform grouping effectively.

### 2.3 Grouping by DSR(Direct Subsumption Relation)

In this section, we suggest a scheme of the grouping which mutually relates the conceptual objects by DSR. This scheme uses the automatic insertion into the existing group when a new conceptual object is applied. If all conceptual object instances of group  $G_2$  satisfy the properties of group  $G_1$ , then the group  $G_1$  subsumes the group  $G_2$ , namely,  $G_2 \subset G_1$ .

The subsumption relation is determined by fuzzy value which describes the properties of conceptual objects.

*(Definition 3) If all super group of  $G_i$  subsumes subgroup  $G_j$  and the property restriction of  $G_i$  subsumes the property restriction of  $G_j$ , then  $G_i$  subsumes  $G_j$ . If the group  $G_i$  subsumes the group  $G_j$ ,  $SR(\text{Subsumption Relation})(G_i \subset G_j)=\{T, F\}$ . In that relation, T means that the subsumption relation is satisfied and F means that the subsumption relation is not satisfied.*

*(Definition 4) If the (Definition 3) is satisfied and there is DSR where one element of subgroup  $G_j$  refers to supergroup  $G_i$  directly which satisfies  $MC(S) \geq MGC$  in  $MC(S)$  then  $DSR\{(c_i, \mu), G_i \geq MGC\} = \{c_i \in G_i \mid G_i \text{ subsumes } c_i\}$  is  $DSR((c_i, \mu), G)=\{(c_i, \mu) \subset G\}$ . Here,  $c_i$  is a conceptual object instance of the group  $G_i$ , and  $G_i$  is a group which satisfies  $MC(S) \geq MGC$  in  $MC(S)$ .*

Now, to subsume a new conceptual object instance, if necessary, an existing group must be defined and a group must be expanded and reconstructed incrementally. To subsume a conceptual object instance, it is necessary to discriminate the existing group which satisfies this conceptual object instance.

For example, if we suppose that there is a group "dog" and conceptual object instances of poodle, jindo-dog satisfies all the restrictions of "dog", subsumption  $\{(a \text{ poodle}, a \text{ jindo-dog}), a \text{ dog}\}=\{T\}$ .

As we can see here, if there exists a structural similarity between instances, the grouping can be created. In other words, subsumption relation  $\{(a \text{ poodle}, a \text{ jindo-dog}), a \text{ dog}\}$  is made of identical properties among the members of a group that (a poodle, a jindo-dog) have. That is the process which the group "dog" is generated using the structured similarity between "poodle" and "jindo-dog".

(Fig. 2) shows an example of grouping by DSR.

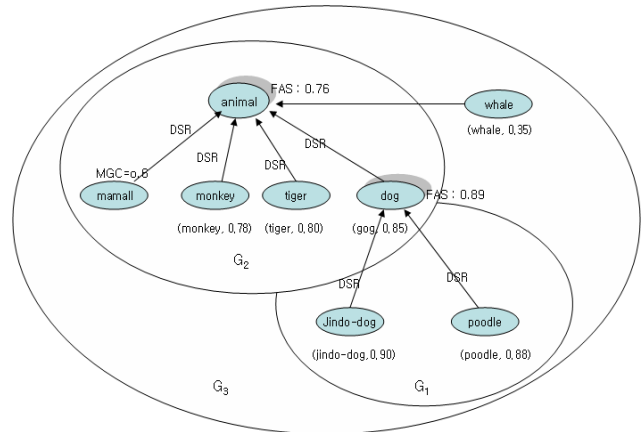


Fig. 2 Conceptual object grouping by DSR

If we perform DSR about  $G_2$  by (Definition 3), (Definition 4).

- ①  $DSR\{(mammal, 0.6), animal\} = \{(mammal, 0.6) \subset animal\}$
- ②  $DSR\{(monkey, 0.78), animal\} = \{(monkey, 0.78) \subset animal\}$
- ③  $DSR\{(tiger, 0.80), animal\} = \{(tiger, 0.80) \subset animal\}$
- ④  $DSR\{(dog, 0.85), animal\} = \{(dog, 0.85) \subset animal\}$ , and if we perform DSR about  $G_1$
- ⑤  $DSR\{(jindo-dog, 0.90), dog\} = \{(jindo-dog, 0.90) \subset dog\}$
- ⑥  $DSR\{(poodle, 0.88), dog\} = \{(poodle, 0.88) \subset dog\}$ . If  $G_1, G_2$  can establish relation of DSF(Direct Subsumption Function) about  $G_3$ ,
- ⑦  $DSR\{\{G_1, G_2\}, G_3\} = \{\{G_1, G_2\} \subset G_3\}$

But if the constraint="zoo" is given, the group "dog" is subsumed by the group "animal" but it is not subsumed by the group that is related to the zoo. Consequently, as this has the effect of decreasing the retrieval number of the group, the efficiency can be improved.

### 3. Experimental Results

This paper evaluated the efficiency of grouping method by using 1000 conceptual objects of animals. For the evaluation of the suggested method, we used SQL server 7.0, Microsoft Visual C++6.0 and MFC of Windows 2000. The depth of group paths is limited as 3 to 6. Too small depth of group paths results too many related conceptual objects. Similarly too large depth of group paths results few corresponding conceptual objects. Retrieval time is also an element of evaluation criteria related directly to the users. The retrieval time is the period between the starting of queries and obtaining the results of retrieval. This paper measured the total average retrieval time of conceptual objects of a group. All the conceptual objects of the same group are assumed to be stored in same file. To inspect the retrieval efficiency of the proposed method,  $CO_{searchtime} = CO_T(1+1/N) + \{(CO_T/group G_i)\} \times O_{Randomkey}$  is used. Table 3. shows the operators for retrieval efficiency.

Table 3. The operators for retrieval efficiency

operators	Meaning
$CO_N$	The number of conceptual objects.
$CO_{time}$	The time spent in finding corresponding conceptual objects which satisfy $MC(S) \geq \mu$ in group $G_i$ .
$GCO_{time}$	The time spent in finding conceptual objects which satisfies FAS in group $G_i$ .
$DIST_{time}$	The time spent in moving group to group, directory to directory, file to group, and directory to file.

Consequently, if there are N groups in the document domain, then the time spent in finding corresponding conceptual objects is  $CO_{time} = GCO_{time} + DIST_{time} \times CO_N$ , and the time spent in searching conceptual objects which satisfy FAS in group  $G_i$  is  $GCO_{time} = x(CO_i) + y(CO_i)$ . Here,  $x(CO_i)$  means total average time to find conceptual objects which satisfies  $MC(S) \geq \mu$  in group  $G_i$  and  $y(CO_i)$  means the total average time to access the related conceptual objects which satisfy FAS in group  $G_i$ .

This paper performed an experimentation by generating random data for an efficiency analysis of the suggested method. The average retrieval time is the average of total time to perform experiment for corresponding conceptual objects of group  $G_i$ .

To inspect the efficiency of the proposed method, 1,000 animal conceptual objects are used in the proposed structure. The experiments are performed 10 times for conceptual objects which satisfy  $MC(S) \geq 0.6$ -cut in group  $G_i$ . The average time is used for comparison. The maximum group degree is limited to 5. Fig.3 shows the experimental result of cases  $MC(S) \geq 0.5$ -cut and  $MC(S) \geq 0.8$ -cut .

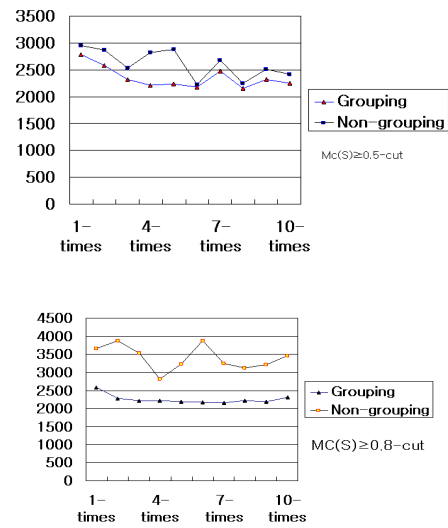


Fig. 3 The results of experiment which satisfies  $MC(S) \geq 0.5$ -cut and  $MC(S) \geq 0.8$ -cut .

Fig.3 shows the average retrieval time of 10 experiments for conceptual object groups. According to the experimental result for  $MC(S) \geq 0.5$ -cut and  $MC(S) \geq 0.8$ -cut, the grouping method is superior to the non-grouping method. In retrieval efficiency the grouping method reduced the retrieval time up to half of that for the non-grouping method. Thus, the grouping method proposed in this paper is superior to the non-grouping method.

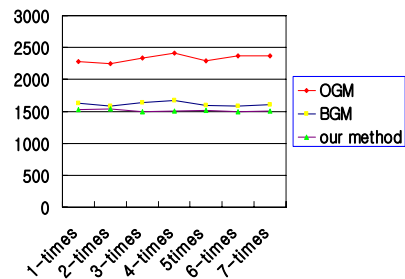


Fig. 4 Total Average Retrieval Time

Figure 4 shows and compares the retrieval time for OGM(Optimistic Genealogy Method)[12], BGM(Balanced Genealogy Method)[12] and the proposed method. The OGM calculates the similarity which is the weighted average among elements in a group. The BGM modified the OGM to solve similarity redundancy. This method uses the parameter  $0 \leq \beta \leq 1$  to calculate the similarity.

The simulation is performed 7 times for 1,000 objects to calculate the total retrieval time for OGM, BGM and the proposed method. According to the simulation result in Figure 4, the proposed method spent less time than the other two methods of OGM and BGM. The simulation result shows that the total retrieval time for the proposed method is 1,514 while those for OGM and BGM are 2,329 and 1,614, respectively. The reason for the difference in the retrieval time is that the

proposed method retrieves the relevant object by FAS and DSR whereas the other two methods retrieve using keywords only. The proposed method has better retrieval performance and is easier to access multimedia documents than OGM and BGM.

#### 4. Conclusion

This paper proposed a grouping method of conceptual objects by  $\mu$ -cut, FAS and DSR structure. The proposed grouping method constructs informations automatically using fuzzy-filtering as the increase of the conceptual object of a domain. The proposed method keeps multimedia document information in a document domain systematically and enables users to service the object information of documents effectively.

Existing methods manage multimedia documents regarding the text mainly as the objects. This paper proposed a grouping method that may manage the formal information as well as the informal information. The grouping structure keeps an inheritance of the super group and the sub group according to their relationship and establishes the subsumption relation according to the inheritance. The subsumption relation between conceptual objects is managed by  $\mu$ -cut, FAS and DSR.  $\mu$ -cut is used in order to determine the subsumption relation which is used to group conceptual objects.

#### References

- [1] Andrea Rodriguez, and Max J. Egenhofer, "Determining Semantic Similarity among Entity Classes from Different Ontologies," *IEEE Trans. on Multimedia*, vol. 15, pp. 442-456, 2003.
- [2] Elisa Bertino, Jianping Fan, Elena Ferrari, Mohand-Said Hacid, Ahmed K. Elmagarmid, and Xingquan Zhu, "A Hierarchical Access Control Model for Video Database Systems," *ACM Trans. on Information Systems*, vol. 21, pp. 155-191, 2003.
- [3] Elina Megalou, and Thanasis Hadzilacos, "Semantic Abstractions in the Multimedia Domain," *IEEE Trans. on Knowledge and Data Engineering*, vol. 15, no. 1, pp. 136-160, 2003.
- [4] Filip Perich, Anupam Joshi, Senior Member, IEEE, Timothy Finin, and Yelena Yesha, Senior Member, IEEE, "On Data Management in Pervasive Computing Environment," *IEEE Trans. on Knowledge and Data Engineering*, vol. 16, no. 5, pp. 621-633, 2004.
- [5] Guojun Lu, "Techniques and Data Structure for Efficient Multimedia Retrieval Based on Similarity," *IEEE Trans. on Multimedia*, vol. 4, no. 3, pp. 372-384, 2002.
- [6] Hisham AI-Mubaid and Syed A. Umair, "A new text Categorization Technique Using Distributional Clustering and Learning Logic," *IEEE Trans. on Knowledge and Data Engineering*, vol. 18, no. 9, pp. 1156-1165, 2006.
- [7] K. Bohms, and T. C. Rakow, "Metadata for Multimedia Documents," *SIGMOD Record*, vol. 23, no. 4, pp. 21-26, 1994.
- [8] Laszlo T. Koczy and T. Geodeon, "Information Retrieval by Fuzzy Relations and Hierarchical Co-occurrence," *Part I. TR97-01, Dept. of Info. Eng., School of Com. Sci. & Eng., UNSW*, pp. 1-18, 1997.
- [9] Prasanna Ganesan, Hector Garcia-Molina, and Jennifer Widom, "Exploiting Hierarchical Domain Structure to Compute Similarity," *ACM Trans. on Information Systems*, vol. 21, pp. 64-93, 2003.
- [10] Yinghui Yang and Balaji Padmanabhan, "GHIC: A Hierarchical Pattern-Based Clustering Algorithm for Grouping Web Transaction," *IEEE Trans. on Knowledge and Data Engineering*, vol. 17, no. 9, pp. 1300-1304, 2005.



**Chongdeuk Lee**

He received the M.S. and the Ph.D. degrees in the Department of Computer Science, Chonbuk National University, in 1989 and 1998, respectively. From 2002 to present, he is a professor at the Division of Applied systems Engineering, Chonbuk National University.



**Taegwon Jeong**

He received the B.S and M.S degrees in Seoul National University, and the Ph.D degree in the University of Florida, all at the Dept. of EE. From 2002 to present, he is a professor at the Division of Applied systems Engineering, Chonbuk National University.