Similarity-based Retrieval of Temporal Documents

Prakash Yamuna K. Selçuk Candan
Computer Science and Engineering Department,
Arizona State University
{prakash.yamuna,candan}@asu.edu

ABSTRACT

In this paper, we describe a similarity-based retrieval framework that addresses the challenges associated with the temporal nature of multimedia documents. Multimedia documents consist of multiple media objects and a set of specifications (eg. temporal) that tie these objects together. Therefore, we describe similarity/dissimilarity measures that aim to capture document authors' intensions. We use a *prioritized* constraint-based framework to evaluate these measures. We also develop algorithms that efficiently compute these measures for special cases.

1. INTRODUCTION

Similarity-based retrieval of temporal information have many applications, most significant of being the multimedia document retrieval. Below, we list two of these applications.

Application I: Multimedia Databases. Multimedia systems employ similarity based retrieval techniques to retrieve images, video, and other media objects. We view a multimedia document as a collection of media objects, which are interlinked with each other through various structures including temporal, spatial, interaction, and user interaction structures. Therefore, in order to be able to retrieve multimedia documents such as SMIL documents the way we retrieve individual media objects, we need to develop similarity-based retrieval techniques that apply to these structures. In this paper, we define the concept of *temporal document similarity.* Note that we are not aware of any other work which defines the concept of *temporal document similarity.*

The structural information, not only enables users to pose structure related queries, but many essential functionalities, such as object prefetching for interactive document visualization, result summarization/visualization, and query processing for document retrieval, depend on the (1) efficiency in representing structural information (2) speed in comparing two documents using their structures, and (3) capability of providing a meaningful similarity value.

Application II: Schedule Databases In order to improve their efficiency, many companies have to deal with temporal scheduling problems. In a recent work, Adalı et al. [2] propose plan databases to manage plans. We believe that, since many scheduling tasks are potentially similar to each other, such a database should also support similarity based retrieval. For instance, given a task to schedule, a user may want to find already created schedules that he or she can use as a starting point that can be tailored to the new task.

In order to address the needs of such applications, we develop distance measures that consider authors' intentions and develop algorithms to efficiently compute them.

ACM Multimedia Workshop Marina Del Rey CA USA

2. BACKGROUND

Figure 1 depicts an overview of the *Integrated Multimedia Management System(IMMS)* architecture, being developed at ASU, which consists of two tightly coupled systems: (1) Media-object Integration System (MIS) and (2) Multimedia Object/Document Base (MODB). MIS is responsible of authoring and presentation of multimedia documents. It integrates various temporal data models, including prioritized instant-based temporal principles we introduced in CHIMP [7, 8]. In order to facilitate various essential tasks, including document indexing, MIS extends these temporal models to enable document scaling and summarization. MODB, on the other hand, is responsible of the storage and retrieval of media objects and multimedia documents.

2.1 Temporal Document Specification

The most basic model that addresses the temporal needs of multimedia applications is the *timeline model.* In this model, the user places events and actions on a timeline. A more flexible formalism is proposed by Allen [4]. In this formalism Allen provides thirteen qualitative temporal relationships (such as *before, meets,* and *overlaps*) that can hold between two intervals. Several other researchers, including *ourselves* in the CHIMP project [8], Buchanan and Zellweger [6], and Song *et.al* [10], on the other hand, proposed the use of a highly-structured class of linear constraints called *difference constraints,* based on the instant-based point formalism introduced by Vilain and Kautz [13]. Constraint-based approaches for document authoring and presentation include Özsoyoğlu *et.al* [9] and Vazirgiannis *et al.* [11].

2.2 Challenges in developing Similarity measures for Temporal Structures

Traditional approaches to retrieval based on dyanmic (such as temporal) properties have been limited to exact matches

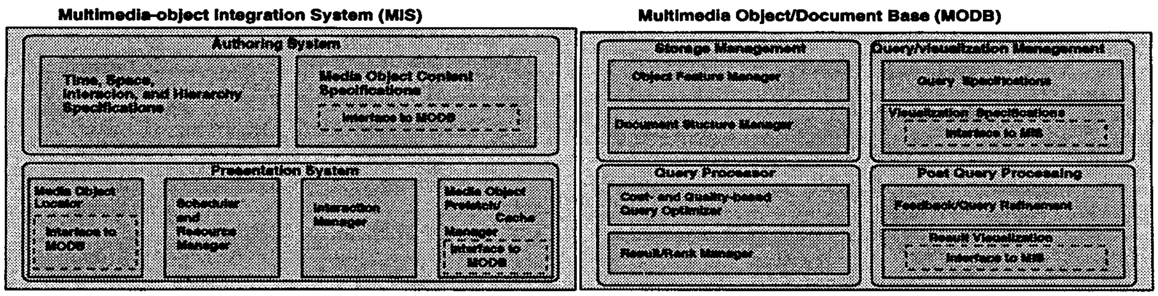

Figure 1: IMMS architecture. Major interfaces between components are denoted by dashed borders.

(e.g., Adalı [1], Del Bimbo [5], Little [3], Vazirgiannis [12]). However, for most retrieval applications, similarity-based retrieval is essential.

The multitude of models by themselves pose a challenge in building a similarity framework for similarity based retrieval. Even if we limit our interest to a single model, such as OCPN, challenges are many. Figure 2 illustrates the challenges assosciated with finding a similarity measure for temporal structures. Some of these challenges include the treatment of number of objects in each document, the temporal structure of the document, the quantitative nature of the temporal relationships, the content of the objects and their role in temporal similarity. In addition, the capabilities of models in supporting synchronization primitives to varying degrees between objects as opposed to a timeline view gives rise to an intensional an extensional view of temporal multimedia documents. Also some synchronization relationships are qualitative in nature (ex. $before(x, y)$), while other have both qualitative and quantitative aspects (ex. $<(x, y, c)$, where c is a positive constant).

3. TEMPORAL DOCUMENT SIMILARITY

In most non-trivial models, dynamic properties are declared as relationships between the media objects. Therefore, a general model must enable us to compare two documents based on the declared *intentions* of the document authors. As a declaration mechanism, constraints have been used by various researchers, including ourselves, to describe the temporal information. We can define a temporal constraint satisfaction problem (TCSP) as follows:

Given a 4-tuple $\langle \mathcal{C}, \mathcal{I}, \mathcal{E}, \mathcal{P} \rangle$, such that

- $\mathcal{C} = \{C_1, \dots\}$ is an infinite set of temporal constants,
- $\mathcal{I} = \{I_1, \dots, I_i\}$ is a set of interval variables,
- $\mathcal{E} = \{E_1, \dots, E_e\}$ is a set of event variables,
- $\mathcal{P} = \{P_1, \dots, P_p\}$ is a set of predicates, where each P_i takes a set of intervals from $\mathcal{I}$, a set of events from $\mathcal{E}$, and a set of *constants* from C, and evaluates to *true* or *false*,

we can define a temporal constraint satisfaction problem as a conjunctive normal formula (CNF) over these predicates, variables, and constants.

Intensional Similarity. According to this general definition of TCSP, each predicate evaluates to either *true* or *false*; and a problem is satisfiable only if the corresponding formula (i.e., each disjunct in the formula) evaluates to *true*. If there are multiple assignments that satisfy the TCSP, then the document has, not one, but a set of solutions. Without a loss of generality, we can assume that that this set is always finite. Figure 3(a) shows the solutions sets of two documents, Doc_1 and Doc_2. Here, C is the set of solutions that satisfy both documents, whereas A and B are the sets of solutions that belong to only one of the documents. We can define the *intensional* similarity of the documents Doc_1 and Doc_2 as $similarity(Doc_1, Doc_2) = \frac{|C|}{|A|+|B|+|C|}$.

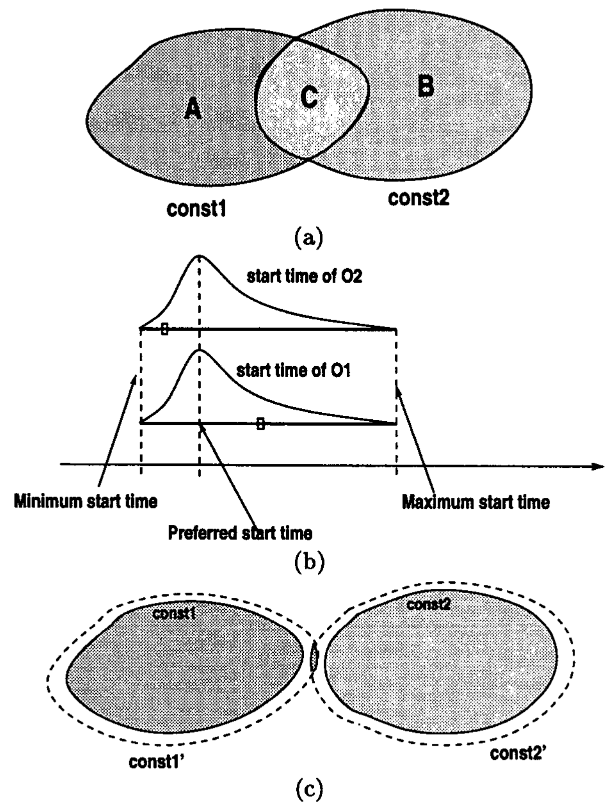

Figure 3: Different models of similarity

Extensional Similarity. However, the above model may fail to capture the *extensional* semantics of the documents: Let us assume that the two lines on Figure 3(b) show the probability distribution of the start time of two objects. By the intensional definition, since the solution sets are identical, the start times of these objects would have a perfect match. However, although their behaviors are identical, there is a chance that, when presented to the user, these two objects will not start at the same time. Therefore, although *intentionally* speaking, the distance between the start times

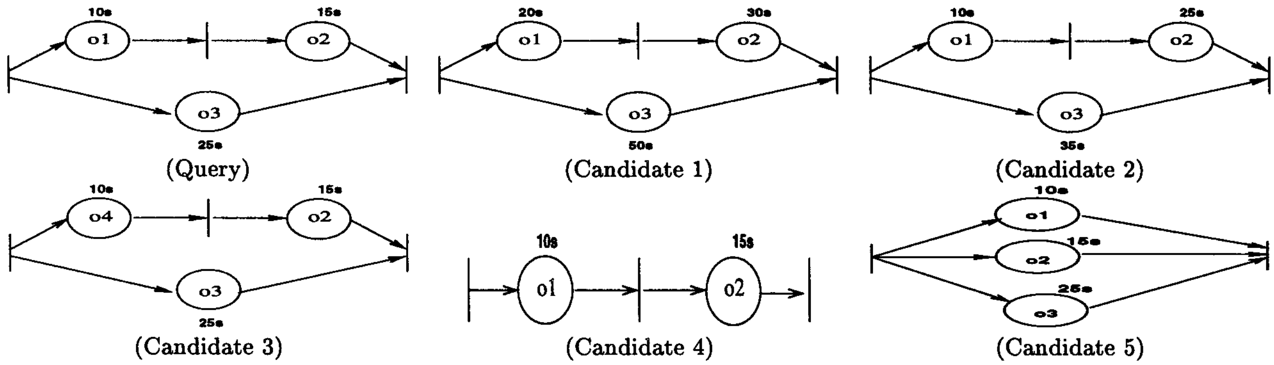

Figure 2: (a) An OCPN graph as query document. Which of the five candidates is most similar to the query?

should be 0, *extensionally* the difference must be non-0. This means that, even when a flexible document is compared with itself, the extensional document distance may be non-zero.

Intensional *Dis*similarity. The intensional semantics given previously, however, has other shortcomings: if an inflexible model (such as the very popular timeline model) is used, then since there is only one solution for a given set of constraints, $\frac{|C|}{|A+B+C|}$ will either evaluate to 1 or 0; i.e., two documents will either match perfectly or will not match at all. It is clear that, such a definition is not useful for *similarity*-based retrieval. A complementary notion of similarity (depicted in Figure 3(c)), on the other hand, captures this case more easily:

- Let us assume that two document Doc_1 and Doc_2 are consistent. Then, since there exists at least one common solution, these documents are similar to each other (*similarity* = 1.0).
- If the solution spaces of these two documents are disjoint, then we can modify (edit) the constraints of these two documents until their solution sets overlap.

We can define the *dissimilarity* between these two documents as the minimum increase required in the sizes of the solution sets for the documents to have a common solution: $dissimilarity(Doc_1, Doc_2) = (|A'|+|B'|)-(|A|+|B|)$, where A' and B' are the new solution sets.

Syntactical Dissimilarity. Note that the two intensional measures given above are complementary: one captures *the degree of similarity between consistent documents* an the other captures *the degree of dissimilarity between in-consistent documents.* In fact, these two measures are (somewhat) related to the popular *edit distance* concept, which defines the dissimilarity between two strings as the number of changes necessary to make the two strings identical. Note that the similarity and dissimilarity measures described above are *semantical*, whereas edit distance is *syntactical*: it can capture the syntactical relationship between the strings `house` and `mouse`, but not the semantical relationship between the strings, *house* and *apartment*. Although the sequence of characters used in a string gives only little clues about the semantics of the string, the set of constraints do actually define the semantics of the document. For instance, two seemingly different sets of constraints can have the same solution set; hence the dissimilarity of the constraints used by document authors does not represent the dissimilarity of the documents. Note, on the other hand, since two documents that have almost the same constraints are likely to have similar contents, it is still possible to extract useful information, more cheaply, from syntactical comparisons.

3.1 Calculating Dissimilarity

In [7, 8] we introduced a prioritized difference constraint based framework to describe temporal properties of multimedia documents. Unlike earlier constraint-based models, in this model each temporal constraint has an associated priority, p, which is a positive number (the higher is the priority, the higher is the relative importance of the constraint). In this paper, we build on this prioritized constraint-based framework as a generic framework that can be used in comparing documents described in most popular models. Note that most conjunctive binary models can easily be described with difference constraints. N-ary models can also be represented by introducing some redundancy. It is also possible to represent disjunctive models with some information loss.

Given a document, D, the corresponding set of difference constraints, $C(D)$, describe the *intension* of its author: Therefore, if two documents, $D1$ and $D2$ are identical or if they represent non-conflicting intensions of their authors, then when the two sets, $C(D_1)$ and $C(D_2)$, of constraints describing these intentions are put together, the resulting set of constraints should not contain any conflicts; i.e., it should be satisfiable. Given two documents, D_1 and D_2, corresponding temporal constraints, $C(D_1)$ and $C(D_2)$, and a mapping μ between the objects in these two documents, we create a merged document, $D_{(1,2)}$. Intuitively, the document merging process identifies those objects that are mapped to each other (i.e., that are identified to be identical by certain process) and merges the constraints in both documents corresponding to such mapped objects. For any unmapped object, it assumes that it is mapped to a 0 length object in the other document, and hence it creates an additional set of constraints that describes this fact. Note that, the above observation gives rise to a simple definition for a distance formula:

DEFINITION 3.1 (SIMPLE DOCUMENT DISTANCE). *The simple distance between two documents, D_1 and D_2, is defined as*

$$\Delta_s(D_1, D_2)_\mu = total\_number\_of\_conflicts\_in(C(D_{(1,2)})).$$

This definition of the document distance is semantically very simple and clean. The triangular inequality, however, does not necessarily hold:

- $\Delta(D_1, D_2) + \Delta(D_2, D_3) \geq \Delta(D_1, D_3)$ **is false.**

This result implies that we can not use spatial index structures or nearest neighbor searches for retrieval of temporal documents. However, there are index structures and clustering algorithms that can be used for retrieval based on a distance measure not satisfying the triangular inequality.

The real disadvantage of this measure, however, is that it is very expensive to compute. It can be shown [8] that in the worst case, the number of conflicts in a document is exponential to the size of the document (in terms of objects and constraints). Therefore, this definition is not very practical. Therefore, we need an alternative measure which is easier to compute.

In [8], we showed that conflicts in temporal specifications show themselves as negative weighted cycles in difference constraint graphs. Therefore, it is possible to identify a set of edges/constraints whose removal would eliminate all conflicts. The minimum number of constraints that needs to be removed to achieve consistency, on the other hand, provides an indication about the reasons of conflicts.

DEFINITION 3.2 (DOCUMENT DISTANCE). *The distance between two documents, D_1 and D_2, is defined as*

$$\Delta(D_1, D_2)_\mu = min\_num\_of\_constraints\_removed(C(D_{(1,2)})).$$

Finding the minimum set of edges to remove is NP-hard. In [8], in order to resolve conflicts during multimedia document presentations, we introduced priorities and used these priorities to approximate the minimality requirement in polynomial time. Therefore, we can define a generic distance measure that captures different relaxation alternatives:

DEFINITION 3.3 (Θ-DOCUMENT DISTANCE). *Given a condition (such as priority-optimality) Θ, the Θ-distance between two documents, D_1 and D_2, is defined as*

$$\Delta_\Theta(D_1, D_2)_\mu = \Theta\_relaxation\_measure(C(D_{(1,2)})),$$

Note, however, the significance of a removed constraint may be different depending on the size of the documents involved in the comparison. For a document with hundred constraints, removal of one constraint may not be significant; yet, for a document with with only a few constraints, removal of a single constraint may be relatively important. Therefore, we need to define a relative document distance as follows:

DEFINITION 3.4 (RELATIVE Θ-DOCUMENT DISTANCE). *Given a condition (such as priority-optimality) Θ, the relative Θ-distance between two documents, D_1 and D_2, is defined as*

$$\Delta_\Theta(D_1, D_2)_\mu = \frac{\Theta\_relaxation\_measure(C(D_{(1,2)}))}{non\_relaxed\_measure(C(D_{(1,2)}))},$$

Note that $\Delta_\Theta(D_1, D_2)_\mu$ will always be between 0.0 and 1.0.

4. CONCLUSIONS

In this paper, we described a similarity-based retrieval framework that addresses the challenges associated with the temporal nature of multimedia documents and temporal schedules. We described a system where users can query, not only the textual content of the media documents, but also the media objects contained within them as well as their temporal structures. For this purpose, we describe intuitive distance measures, which consider document authors' intentions, applicable for different retrieval tasks and we develop algorithms that efficiently compute some of these measures. Future work will include development of a benchmark database and investigation of efficient approximations for other models, as well as techniques for indexing temporal structures.

5. REFERENCES

[1] S. Adalı, M.-L. Sapino, and V. Subrahmanian. A multimedia presentation algebra. In *SIGMOD'99*, pages 121–132, June 1999.

[2] S. Adalı, M. L. Sapino, and V. S. Subrahmanian. Plan databases: Model and algebra, 2000.

[3] G. Ahanger and T. Little. Data semantics for improving retrieval performance of digital news video systems. In *DS-8 1999*, pages 47–64, 1999.

[4] J. F. Allen. Maintaining knowledge about temporal intervals. *Communications of the ACM*, 26(11):832–843, November 1983.

[5] A. D. Bimbo, E. Vicario, and D. Zingoni. Symbolic description and visual querying of image sequences using spatio-temporal logic. In *IEEE Transactions on Knowledge and Data Engineering, Vol. 7, n.4*, August 1995.

[6] M. C. Buchanan and P. T. Zellweger. Scheduling multimedia documents using temporal constraints. In *Network and Operating System Support for Digital Audio and Video*, pages 237–249, 1992.

[7] K. S. Candan, B. Prabhakaran, and V. S. Subrahmanian. Chimp: A framework for supporting distributed multimedia document authoring and presentation. In *Proceedings of ACM Multimedia'96*, pages 329–340, 1996.

[8] K. S. Candan, B. Prabhakaran, and V. S. Subrahmanian. Collaborative multimedia documents: Authoring and presentation. *International Journal of Intelligent Systems on Multimedia Computing Systems*, 13(12), 1998.

[9] V. Hakkoymaz, J. Kraft, and G. Ozsoyoglu. Constraint-based automation of multimedia presentation assembly. *Multimedia Systems*, 7(6):500–518, 1999.

[10] J. Song, Y. N. Doganata, M. Y. Kim, and A. N. Tantawi. Modeling timed user-interactions in multimedia documents. In *Proceedings of the International Conference on Multimedia Computing and Systems '96*, pages 407–416. IEEE Computer Society, 1996.

[11] M. Vazirgiannis and S. Boll. Events in interactive multimedia applications: Modelling and implementation design. In *Proceedings of the International Conference on Multimedia Computing and Systems*, pages 244–251. IEEE Computer Society, 1997.

[12] M. Vazirgiannis, Y. Theodoridis, and T. Sellis. Spatio-temporal composition and indexing for large multimedia applications. In *Multimedia Systems 6(4)*, pages 284–298, 1998.

[13] M. Vilain and H. Kautz. Constraint propagation algorithms for temporal reasoning. In *Proceedings of AAAI'86 Conference on Artificial Intelligence*, pages 377–382, 1986.