# Integrating XLink and XPath to Retrieve Structured Multimedia Documents in Digital Libraries

**Zhigang Kong & Mounia Lalmas**
Department of Computer Science,
Queen Mary, University of London
Mile End Road, E1 4NS London, UK
{cskzg, mounia}@dcs.qmul.ac.uk

**Abstract**

To support the retrieval of multimedia data according to user information needs, multimedia information retrieval in digital libraries must be based on semantics and not just primitive features of multimedia objects. In this paper, we investigate the intersection of the hierarchical and linking structural information with the content information to support the retrieval of multimedia data in digital libraries based on XML format. For this purpose, we integrate XLink and XPath information to exploit the intersection of hierarchical and linking information within XML-based multimedia documents.

Keywords: semantic multimedia retrieval, XPath, XLink, digital library.

## 1. Introduction.

Multimedia information systems are widely recognised to be one of the most promising fields in the area of information management (Baeza-Yates & Ribeiro-Neto, 1999). Digital libraries are very complex multimedia information systems whose building process is both expensive and resource-intensive (McCray & Gallagher, 2001). One of the most important characteristics of digital libraries is the variety of data they must be able to support. Digital libraries must have the capability to store, retrieve, transport, and present data with heterogeneous characteristics such as text, images, videos, graphics and sound. For this reason, the development of a digital library is considerably more complex than that of a traditional information management system. Conventional systems only deal with simple data types, such as strings or integers. On the contrary, digital libraries support multimedia objects often embedded in a complex structure. Therefore, multimedia information retrieval that considers structure information is essential for the success of digital libraries.

As structured information is increasingly encoded using XML, multimedia objects in digital libraries can and will be more and more embedded in XML documents. *The question is how to effectively retrieve XML-based multimedia documents*. As a kind of hypermedia with controlled structure, XML-based multimedia documents in digital libraries have a strict hierarchical structure as well as a linking structure. The structure of these multimedia documents can play an essential role in the semantic-based multimedia information retrieval in digital libraries. It is well known (Baeza-Yates & Ribeiro-Neto, 1999; Schatz, 1997) that to satisfy user information needs, multimedia information retrieval in digital libraries must be based on semantics and not just primitive features of multimedia objects. Therefore, integrating the hierarchical and linking structure information with the content information seems fundamental to the effective retrieval of multimedia data in XML documents. It is such an approach that we propose and investigate in this paper.

More specifically, we investigate an approach to XML-based multimedia documents that exploits XPath and XLinks. XPath models hierarchical structure trees of XML document nodes and XLink presents an explicit relationship between resources or portions of resources in XML documents. The aim of this approach is to integrate the hierarchical and linking structure information with the content information to support the retrieval of multimedia data in XML-based digital libraries.

This paper is organised as follows. In the next section (section 2) we provide background and related work information. In section 3, we describe our approach to represent the content of a given multimedia object. We define the object domain knowledge for the given multimedia object as composed of regional knowledge and linking knowledge. In section 4, we define the regional knowledge that we use to represent the multimedia object. The regional knowledge allows for text-based retrieval of multimedia and hence any text-based information retrieval model can be used. In section 5, we describe the linking knowledge and exploit the liking information based on XLink issues. In section 6, we work through an example of how retrieval would work. In the last section we conclude and discuss future work.

## 2. Background and Related Work

Digital libraries are very complex multimedia information systems whose building process is both expensive and resource-intensive (McCray & Gallagher, 2001). The underlying data model, the query language, and the access and storage mechanisms of digital libraries must deal with the networked collections of digital text, documents, images, sounds, scientific data, software etc, that is, multimedia data (Reddy & Wladawsky-Berger, 2001). Therefore, effective multimedia information retrieval is essential for the success of digital libraries.

The primary purpose of digital libraries is to enable searching of electronic collections distributed across networks, rather than merely creating electronic repositories from digitised physical materials (Schatz, 1997). Also, the benefits of digital libraries will not be appreciated unless they are easy to use effectively (Lynch & Garcia-Molina, 1995). Digital Libraries are constructed – collected and organised – by a community of users. Their functional capabilities support the information needs and uses of that community (Baeza-Yates & Ribeiro-Neto, 1999). To support the retrieval of multimedia data according to user information needs, multimedia information retrieval in digital libraries must be based on semantics and not just primitive features of multimedia objects. In this paper, we concentrate on the semantic retrieval of multimedia data in digital libraries.

Semantics based retrieval of multimedia data in digital libraries can be based on the "hypermedia" nature of most digital libraries. Dunlop (Dunlop, 1991; Dunlop, 1993) introduced a model for multimedia retrieval in a hypermedia system in which users could issue a textual query first to obtain one or more textual nodes and then navigate from these text nodes to the desired non-textual (multimedia) node. The textual nodes linked to the multimedia (non-textual) node were used to define an artificial descriptor of the data that permits direct retrieval of it from a text-based query. Based on Dunlop's model, Harmandas et al (Harmandas et al, 1997) apply this technique to investigate the retrieval of images on the World Wide Web. Our proposed approach benefits from the "hypermedia" nature of digital libraries by integrating the "hyper-linking" information in the knowledge representation of multimedia object.

Chiaramella et al (Chiaramella et al, 1996a; Chiaramella et al, 1996b) presented an approach to hypermedia systems that integrates hypermedia and IR (Information Retrieval). The approach argued IR and hypermedia as two complementary approaches that must be integrated to provide a more efficient and effective environment for accessing and retrieving information. A model for structured multimedia document retrieval was presented, and its basic principle was that such an integrated model allows the manipulation of the logical structure of documents, the navigation links, the semantic content and the attributes of document components. In addition, Chiaramella et al define the notion of *index objects*, which correspond to classes of structural objects that constitute retrievable objects (Chiaramella et al, 1996b). More precisely, given a hierarchical document structure only nodes of specific types may form the roots of index objects. In our approach, we use the textual index objects to represent the multimedia (non-textual) objects; that is, the retrieval of multimedia objects is based on textual index objects. The *tf-idf* weighting formula in classical IR can then be applied to represent and retrieve the multimedia objects.

The "hypermedia" nature can be combined with low-level feature processing to improve the retrieval of multimedia objects. Sclaroff et al (Sclaroff et al, 1999) provided an approach of combining textual and visual cues for content-based image retrieval on the World Wide Web. The combined approach allows improved performance in content-based search. Further, Henrich and Robbert (Henrich & Robbert, 2000) pointed out that an efficient combination of automatic text retrieval, retrieval in metadata (usually created manually) and content-based retrieval on multimedia objects is needed for searching structured multimedia documents in digital libraries. Our approach uses XLinks to link pairs of similar multimedia objects based on low-level feature analysis. We can then retrieve a multimedia object because it can be inferred from a retrieved multimedia object from the XLinks. We call this process a "relevance inference".

A hypermedia document has two aspects: content and structure. It is believed that the integration of content search with structure analysis can improve hypermedia documents retrieval (Chakrabarti et al, 1998; Gibson et al, 1998). Several approaches have been developed for searching web documents, a "loose" hypermedia system. Page et al (Page et al, 1998) defined PageRank, a global ranking scheme, which tries to capture the notion of "importance" of a page: the page receives more importance when many important pages point to it. Another web retrieval approach is HITS (Hypertext Induced Topic Search) that was first proposed by (Kleinberg, 1999). HITS is a query-dependent ranking technique that produces two ranking scores, the authority score and the hub score between which a mutually reinforcing relationship exists. This has been shown to be effective in web retrieval. We use similar paradigms for retrieving multimedia objects that can be inferred by other retrieved objects.

Multimedia objects in digital libraries will be increasingly embedded in XML documents. Therefore, information retrieval of XML documents plays an important role in digital libraries. Fuhr (Fuhr, 2001; Fuhr, 2002; Fuhr, 2003) presented a document-centric view of XML that takes into account the intrinsic imprecision and vagueness of IR and the resulting need for ranked lists as search results. This approach extends the XPath (Clark et al, 1999) part of an XQuery-like (Boag et al, 2003) query language by the following features: weighting, ranking, and relevance-oriented search. In our paper, we retrieve parts of XML documents, *index objects*, and then use them to form the representation of a multimedia object. After applying the classic IR weighting formula to the *index objects*, we can compute the term frequency of the representation by combining the term frequencies of the corresponding *index objects*.

XPath and XLink play an important role in XML retrieval. XPath is a language whose primary purpose is to address parts of an XML document. In support of this primary purpose, it also supplies basic facilities for manipulation of strings, numbers and Booleans. XPath uses a compact, non-XML syntax to facilitate the use of XPath within URIs (Uniform Resource Identifiers) and XML attribute values. XPath operates on the abstract, logical structure of an XML document (Clark et al, 1999). For example, /article/sec[2]/p[6] selects the sixth paragraph of the second section of the article. XML Linking Language (XLink) is defined by W3C to allow elements to be inserted into XML documents in order to create and describe links between resources. Based on the XML syntax, XLink can be used to create links similar to the simple unidirectional hyperlinks of today's HTML, as well as more sophisticate links (DeRose et al, 2001).

XPath models an XML document as a tree of nodes in which a hierarchical structure exists. XLink presents an explicit relationship between resources or portions of resources in XML documents. Integrating XLink and XPath thus combines hierarchical information and linking information within XML documents. Combination of these two kinds of structure information analysis could give more accurate cues for the semantic retrieval of XML-based multimedia documents. This paper investigates this combination.

## 3. Object Domain Knowledge

Multimedia objects in XML documents are often organised as leaf nodes in a hierarchical structure. Those XML documents usually contain substantial textual nodes. Information about a multimedia object can be found in the textual nodes associated with it. Therefore, the text information within the textual nodes associated with a multimedia node can be used to calculate a representation of it that is capable of supplying direct retrieval of this multimedia data by textual (natural language) query.

In addition, we can find useful information about a given multimedia object from the objects, either text objects or other multimedia objects, linked to it. This linking information can be used to represent the content of this given multimedia object or to provide relevant relationships between this multimedia object and other multimedia objects. Furthermore, the XLinks themselves can provide meaningful information according to their semantics, attributes, arc traversal rules, types of links and linking relationships.

Based on the above ideas, we consider a multimedia object's surrounding hierarchical (text) information and hyperbase linking information as a domain that can supply knowledge that describes the multimedia object's content. Then we define the knowledge involved in the domain of a multimedia object's hierarchical and linking information as its *object domain knowledge*. We further define the knowledge included in a multimedia object's hierarchical inner information as its *regional knowledge* and that involved in its hyperbase linking information as its *linking knowledge*. Therefore, a multimedia object's object domain knowledge is composed of a regional knowledge and a linking knowledge.

XPath can be used to retrieve parts of XML documents, *index objects*, within their hierarchical structure. These index objects form the regional knowledge of multimedia objects. XLinks presents an explicit relationship between hyper-linked index objects that constitute the linking knowledge of multimedia objects. Our aim is to integrate XLink and XPath to exploit the object domain knowledge, the combination of regional knowledge and linking knowledge, to support the effective retrieval of multimedia objects in XML-based digital libraries. How these are used for the retrieval of multimedia objects is discussed in section 4 and 5, respectively.

## 4. Regional knowledge

XML documents are composed of structured elements. IR in XML and in structured documents is different from classic IR. Researchers in these fields suggest treating the structured elements as atomic retrieval units, in the fashion that classic IR models have treated documents (Chiaramella et al 1996b; Fuhr 2002; Fuhr 2003). They define these atomic retrieval units as *index objects*. Then the classic weighting formula (i.e. *tf-idf*) can be applied to the index objects in XML documents.

The textual index objects surrounding a multimedia object based on its hierarchical structure can be used to form its regional knowledge. Then we can calculate the term frequency (*tf*) of any given term in a multimedia object's regional knowledge by combining the *tfs* of the given term in the surrounding index objects.

To calculate the representation of a multimedia object's regional knowledge, we split the text information associated with a multimedia object according to its position with respect to the multimedia object. In XML documents, we identify three types of text information associated with multimedia objects:

1. Caption or Description.
2. Sibling text information.
3. Hierarchical text information.

The definitions for the above three types of text are given below together with examples. The latter are derived from the INEX[1] data. At first, "Caption" or "Description" refers to the text information within, or tightly associated with, multimedia tags, which is normally used to describe multimedia objects. For example, in the INEX document collection one kind of multimedia tags is <fig> (Figure 1, Section 6). Then the Caption can be defined as the text between <fig> and </fig>. Secondly, "Sibling" text information is the text in the index objects that are sibling to the multimedia object. These could be either text objects or multimedia objects. In addition, the objects we called sibling should be in the same hierarchical level as the multimedia object in the logical structure. The sibling objects could be either the two paragraphs just before or following the multimedia object or any number of objects surrounding the multimedia object in the same hierarchical level. For example, the text within the <p> element that is located just before or after <fig> element, or the caption of another <fig> element that is just before or after this <fig> element. At last, hierarchical text information is defined as text in higher-level index objects within the hierarchical structure that can be viewed as knowledge capable of being inherited by the lower level element.

The three types of text defined above may contribute differently in representing the content of a multimedia object. Text closer to a multimedia object can be thought to form a better representation of that object than text less close to the multimedia object. Based on this idea, we assign a parameter $k$ to different types of text to reflect their contribution or, what we call "*relevance strength*" in representing the content of a multimedia object.

For any given multimedia object *o* having regional knowledge composed of *n* index objects *I*, and a term *t*, the term frequency (*tf*) of term *t* in the regional knowledge of given object *o* can be calculated as follow:

$$W(t,o)=\frac{1}{n}\sum_{\forall I} k_I w(t,I) \qquad (1)$$

Here $k_I$ is the *relevance strength* of index object I and *w(t, I)* is the term frequency of term *t* in *I*.

For instance, in the example of section 6, we have 5 index objects for the regional knowledge of I00011.gif (1 Caption, 2 Sibling information, and 2 Hierarchical information). Thus, $n = 5$, and there are five corresponding *w(t, I)* for the five index objects, and each has a relevance strength $k_I$. Then we can calculate the *W(t, o)* according to the above formula.

With formula (1), or any similar formalism, we arrive at a representation of the multimedia object that is exactly the same as the representation of text data. Therefore, we can use any text-based IR model to perform retrieval (e.g. vector space model, language model, etc). What remains is to investigate a number of issues:

- What kind of "sibling" information is more effective? For instance, we can consider two sibling elements, the one before and the one following the multimedia object to form the sibling representation. Alternatively, more sibling elements (i.e. further away from the object itself) can be considered as sibling information.
- Investigate the influence of hierarchical level when forming the hierarchical information representation. For instance, for a multimedia object in \\chapter2\section5[2], the hierarchical information in the section 5 element may form a better representation of the multimedia object than that in chapter 2 element or vice versa.
- Evaluate what set of *relevance strength* values is to be allocated to the three types of text when forming the regional knowledge.

[1] The INEX (Initiative for the Evaluation of XML Retrieval) is part of a large-scale effort to encourage research in information retrieval and digital libraries. It provides a large test collection of XML documents, uniform scoring procedures, and a forum for organisations to compare their results. http://www.is.informatik.uni-duisburg.de/projects/inex/index.html.en

[2] XPath notations are explained later in the paper (see Section 6).

These issues will be investigated empirically, and are the purpose of our future work.

## 5. Linking Knowledge

There are two kinds of information in a multimedia object's linking knowledge: a multimedia object can be either linked to text information or to another multimedia (non-text) object. Textual linking information can be used to provide direct retrieval of the linked multimedia object by natural language (text) query. Although the multimedia linking information cannot provide textual cues for the linked multimedia object, it can still be used to retrieve the linked multimedia object. Due to the different character of these two kinds of linking information, we discuss them separately in the following two subsections.

### 5.1 Textual linking information

Based on XLink rules (DeRose et al, 2001), we can organise sophisticated links between resources, which include index objects as well as other types of data. The resources could be either metadata describing the content of the multimedia, author information or any other information associated with this multimedia object. For example, a picture of the Tower Bridge in London can have five textual resources linked to it: one resource being a description of the content of this image (metadata), another being a description of the picture's author information, two resources providing introduction of two historical events of the bridge, and the last resource describing the building of the bridge.

Suppose that the file TowerBridge.gif (GIF image of the Tower Bridge in London) is stored at http://www.towerbridge.co.uk/TowerBridge.gif, the two files describing historical events are event1.xml and event2.xml, and the file describing the building of this bridge is building.xml. These three files are also stored at http://www.towerbridge.co.uk. Let the picture's author be Peter Hunter whose information is stored in PeterH.xml, and located at http://www.towerbridge.co.uk/people. Let the metadata be in the file pictures.xml, and stored in the same directory of TowerBridge.gif. The metadata format for the TowerBridge.gif in pictures.xml is as follow:

```
…
<pic id = "00066">
    <title>Tower Bridge</title>
    <author>Peter Hunter</author>
    …
</pic>
…
```

Then the XLinks of this example is presented as follow:

```
<extendedlink xlink:type="extended">
  <loc xlink:type="locator" xlink:href="http://www.towerbridge.co.uk/TowerBridge.gif"
       xlink:label="pic66" xlink:title="Tower Bridge"/>
  <loc xlink:type="locator" xlink:href="http://www.towerbridge.co.uk/pictures.xml#xpointer(//pic, [@id = '00066'] )"
       xlink:label="resource1" xlink:title="metadata"/>
  <loc xlink:type="locator"
       xlink:href="http://www.towerbridge.co.uk/pictures.xml#xpointer(string-range(//author, 'Peter Hunter'))"
       xlink:label="author5"/>
  <loc xlink:type="locator" xlink:href=" http://www.towerbridge.co.uk/event1.xml" xlink:label="resource2"/>
  <loc xlink:type="locator" xlink:href="http://www.towerbridge.co.uk/event2.xml" xlink:label="resource3"/>
  <loc xlink:type="locator" xlink:href="http://www.towerbridge.co.uk/building.xml"
       xlink:label="resource4" xlink:title="The building of Tower Bridge"/>
  <loc xlink:type="locator" xlink:href="http://www.towerbridge.co.uk/people/PeterH.xml"
       xlink:label="person7" xink:title= "Photographer Peter Hunter"/>

  <go xlink:type="arc" xlink:from="pic66" xlink:to="resource1" xlink:show="none"/>
  <go xlink:type="arc" xlink:from="resource2" xlink:to="pic66" xlink:show="embed" xlink:actuate="onLoad"/>
  <go xlink:type="arc" xlink:from="resource3" xlink:to="pic66" xlink:show="embed" xlink:actuate="onLoad"/>
  <go xlink:type="arc" xlink:from="resource4" xlink:to="pic66" xlink:show="embed" xlink:actuate="onLoad"/>
  <go xlink:type="arc" xlink:from="author5" xlink:to="person7" xlink:show="new" xlink:actuate="onRequest"/>
</extendedlink>
```

We can integrate this textual information in XLinks to the *regional knowledge* discussed in the previous section to provide additional textual information related to the multimedia object, which can then provide further and more enhanced description of the multimedia object.

For the text in the metadata, text in the attribute, and text in the linked resources, we should use different representations and different models. Queries for meta-data and text in attributes could be database-like queries; the Boolean model would therefore be suitable for these representations. In the previous XLink example, this would concern text in the pictures.xml and xlink:title=”The building of Tower Bridge”. As the “title” is one of the semantic attributes, the text in it can describe the meaning of the resource building.xml within the context of the above link. Retrieval of textual information from linked resources could be based on typical IR queries and *tf-idf* ranking strategies could be used to retrieve them. In the previous example, this would apply to the text in event1.xml, event2.xml, building.xml and PeterH.xml.

Textual information in the different linked resources may contribute differently in describing the content of the multimedia object. In order to combine different parts of text together, we need to calculate the relative strength of their contribution. The following linking structure information affects the contribution, also called *relevance strength*, of the text: types of attributes, arc traversal rules, types of links and linking relationships.

There are several types of attributes in XLinks: Locator Attribute, Semantic Attributes, Behaviour Attributes and Traversal Attributes. Each type of attribute has its own function that could be used to decide the relevance strength. Traversal rules could be also useful in determining the contribution of different linking information. In the previous example: <go xlink:type=”arc” xlink:from=”resource2” xlink:to=”pic66” xlink:show=”embed” xlink:actuate=”onLoad”/>, “show” and “actuate” attributes are behaviour attributes. The former decides that the picture “pic66” is embedded in the “resource2” and the latter means that an application should traverse to the ending resource immediately on loading the starting resource. Further, the links can be divided into different types such as illustration, explanation, and documentation types. Resources in each type of links could have different contributions. At last, linking relationships, which have been exploited by many web search approaches (Page et al, 1998 and Kleinberg, 1999), could also be used to calculate the relevance strength of linking information.

Based on the above ideas, we can use additional related text in the hyper-linking information to constitute the representation of the multimedia objects. By integrating the text linking information and the regional knowledge, the representation could describe the multimedia objects more accurately and provide more effective retrieval. Our proposal is to use XLinks to exploit the textual linking information and then integrate this text in the representation of multimedia objects.

### 5.2 Multimedia linking information

A link between two multimedia objects can be considered as a cue that these two multimedia objects are related to each other in some way. Thus if one of the two multimedia objects is retrieved for a given query q, the other multimedia object can be inferred as relevant to this query q, and thus be retrieved. We call the process of retrieving a multimedia object by inference of a related retrieved multimedia object as a “*relevance inference”*. Three factors affect a *relevance inference* process, and as such, the retrieval status value of the multimedia object that is being “inferred”:

- The retrieval status value of the multimedia object that infers the given one. The relevance strength of a multimedia object depends on the multimedia objects that connect to it by XLinks. In other words, a multimedia object can obtain a higher relevance strength if it has been inferred by a higher ranked retrieved multimedia object than by lower ranked retrieved one.
- The number of multimedia objects that infer the given one. The multimedia objects linked by more objects whose contents are related to the query are more likely to be related to the query when compared to those that are linked by fewer objects.

- The XLinks that connect the given multimedia object. The semantics and attributes within XLinks, types of XLinks and the link relationship between the linked objects can affect the process of *relevance inference*.

In our work, we currently consider only XLinks that are built based on low-level feature analysis. By extracting low-level features of multimedia objects, we can automatically allocate XLinks between pairs of multimedia objects. These XLinks are bidirectional because the objects are similar to each other so we can follow the link from one to the other and vice versa. The XLinks are then stored in a Linkbase, in which the similarity value between pairs of multimedia objects forms the attribute value. We can use a threshold value so that only pairs of multimedia objects whose similarity value is higher than the threshold value are linked together.

As the similarity values between pairs of multimedia objects can be pre-computed off-line, the online computing cost is reduced. Therefore, it is possible, when considering efficiency issues, to combine textual retrieval and content-based retrieval online based on this XLink approach. We first retrieve a set of multimedia objects based on their object domain knowledge and then using the relevance inference, we find all the multimedia objects similar to this initial set.

Based on the XLinks between pairs of multimedia objects, the multimedia objects that have little or no associated text information can still be retrieved if they are linked to the multimedia objects that have enough text information in their object domain knowledge. Therefore, the list of retrieval results obtained from the previously "text-based" retrieval can be enhanced with additional (hopefully) relevant objects. In addition, highly relevant objects will then be highly ranked due to XLinks. As the links are allocated based on low-level feature similarity, this approach provides one possible way to integrate low-level feature process into text query to bridge the semantic gap well known in multimedia retrieval (Gudivada and Raghavan, 1995).

An advantage of this approach is that this kind of XLinks can be automatically computed. As manually allocating XLinks is a time-consuming and error-prone process, this approach provides an easy way to reorganise the collection of multimedia objects by XLinks.

## 6. Example

We use the INEX data as an example to illustrate the object domain knowledge for a multimedia object. INEX is a test collection used for the evaluation of XML retrieval, in the context of digital libraries, and as such provides a realistic scenario of an XML-based multimedia digital library. In the Figure 1, "article" is the root object, "author", "title", "abs", "sec1", "sec2" are child objects of the article. In section 1 (sec1) there are several paragraphs ("P" tags) and one figure ("Fig1" tag). Fig1 is linked to Fig8 and Fig9 (for example because they have similar low-level features). Other XLinks link Fig1 to some remote textual information.

Based on the above idea, the object domain knowledge of Fig1-I00011.gif can be identified as follows. Textual information within the tag <Fig1> can be viewed as the Caption; text in the paragraphs that are either just before or after Fig1 tag ("P1" and "P2") is the Sibling text information; the text within "sec1" and "article" tag can be considered as Hierarchical text information. These three parts form the regional knowledge of I00011.gif. Fig8 and Fig9 are the multimedia linking information of I00011.gif. Other XLinks of I00011.gif link it to some remote textual information that could be either metadata describing this figure or some paragraphs that involve information related to this figure. The integration of all the above is the Object Domain Knowledge of Fig1-I00011.gif.

The XPath expressions of the regional knowledge of the I00011.gif are as following:

- Caption or Description: /article/section [1]/fig [@id="I00011.gif"], -- meaning the 1st section of the article, and with id value I00011.gif.

- Sibling information: /article/section [1]/p [1], --meaning the 1st paragraph, of the 1st section in the article; /article/section [1]/p [2], -- meaning the 2nd paragraph, of the 1st section in the article.
- Hierarchical information: /article, -- meaning the article itself; /article/section [1], -- meaning 1st section of the article.

In addition, textual linking information can be integrated in the regional knowledge as discussed in section 5.1.

Furthermore, Fig 8 and Fig 9 can be retrieved according to *relevance reference* when the Fig 1 is retrieved.

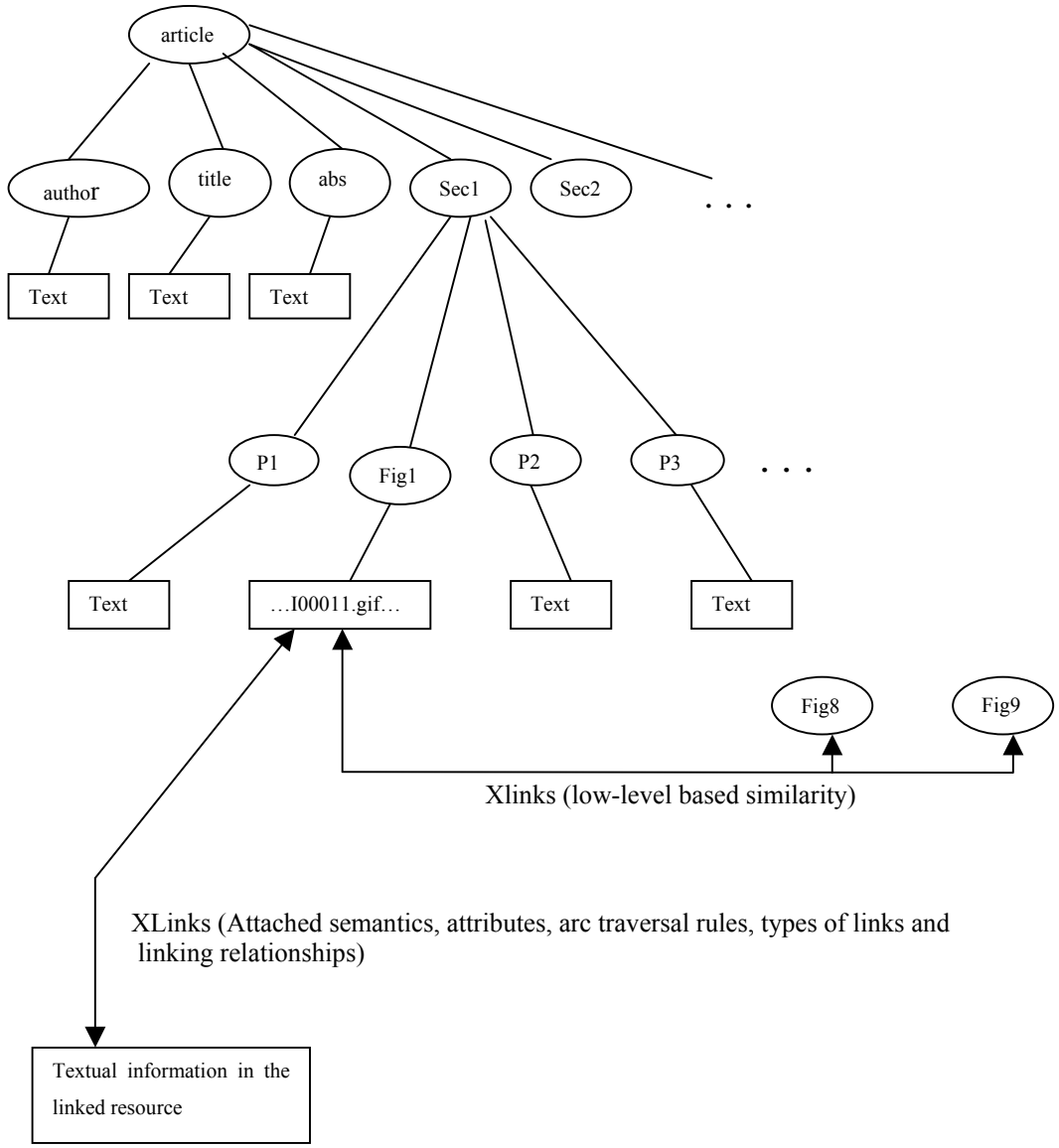


**Figure 1.** Example of an XML-based multimedia document

## 7. Conclusion and Future work

In this paper, we investigate a representation of object domain knowledge that, by exploiting XPath and XLinks, can be used to integrate the hierarchical and linking structural information with the content information to support the retrieval of multimedia objects in digital libraries. We define the

object domain knowledge for the given multimedia object to be composed of a regional knowledge and a linking knowledge. Our approach divides the retrieval process into two steps. At first we can apply any text-based IR model, by exploiting the regional knowledge and the text linking information of multimedia objects, to provide direct retrieval of multimedia objects by natural language (text) queries. Then we integrate the multimedia linking information, which takes into account the low-level feature similarity with previously retrieved results to perform a relevance inference process in order to refine the final results.

Our approach is a first step towards the use of upcoming standards for retrieving in an integrated manner multimedia objects in the context of digital libraries.

Further work will investigate the effectiveness of the representation, the so-called object domain knowledge, based on the INEX collection. The following issues will be investigated: what set of relevance strength values should be allocated to the three types of text when forming the regional knowledge; how to integrate linking structure information and text content to calculate the representation of text linking information; and the effectiveness of the relevance inference process.


## Acknowledgement

We would like to thank Tassos Tombros for his comments.


## References:


Baeza-Yates, R. & Ribeiro-Neto, B. (1999). *Modern Information Retrieval*. Addison Wesley, Essex, England, 1999.

Boag, S.; Chamberlin, D.; Fernández, M.; Florescu, D.; Robie, J.; & Siméon, J. (2003). *XQuery 1.0: An XML Query Language*. http://www.w3.org/TR/xquery/

Chakrabarti, S., Dom, B., & Indyk, P. (1998). Enhanced Hypertext Categorization Using Hyperlinks. In Proceedings of the ACM SIGMOD Conference, 1998.

Chiaramella, Y. & Kheirbek, A. (1996a). *Information Retrieval and Hypertext*, chapter An Integrated model for Hypermedia and Information Retrieval. Kluwer Academic Publ., 1996.

Chiaramella, Y., Mulhem, P., & Fourel, F. (1996b). *A Model for Multimedia Information Retrieval*. Technical report, FERMI ESPRIT BRA 8134, University of Glasgow.

Clark, J. & DeRose, S. (1999). *XML Path Language (XPath) Version 1.0.* http://www.w3.org/TR/xpath

DeRose, S. Maler, & E. Orchard, D. (2001). XML Linking Language (XLink) Version 1.0. http://www.w3.org/TR/2001/REC-xlink-20010627/

Dunlop, M. D. (1991). *Multimedia Information Retrieval*. PhD Thesis. Department of Computer Science, University of Glasgow, Report 1991/R21.

Dunlop, M.D. & Van Rijsbergen C. J. (1993). Hypermedia and free text retrieval. *Information Processing & Management*, 29(3), pp. 287-298.

Fuhr, N. & GroBjohann, K. (2001). XIRQL: A Query Language for Information Retrieval in XML Documents. In: Croft, W.; Harper, D.; Kraft, D.;Zobel, J. (eds.): *Proceedings of the 24th Annual International Conference on Research and development in Information Retrieval,* Pages 172-180. ACM, New York.

Fuhr, N. & Weikum, G. (2002). Classification and Intelligent Search on Information in XML. Bulletin of the IEEE Technical Committee on Data Engineering, 25(1), 2002.

Fuhr, N. (2003). XML Information Retrieval and Information Extraction. In: Franke, F.; Nakhaeizadeh, G.; Renz, I. (eds.): Text Mining. Theoretical Aspects and Applications.

Gibson, D., Kleinberg, J., & Raghavan, P. (1998). Clustering categorical data: an approach based on dynamical systems. In Proc. VLDB'98, New York, NY, 1998.

Gudivada, V. N. & Raghavan, V. V. (1995). Guest editors' introduction: Content-based image retrieval systems. Computer, 28(9): 18-22.

Harmandas, V., Sanderson, M., & Dunlop, M.D. (1997). *Image retrieval by hypertext links*. Proceedings of SIGIR-97, 20th ACM International Conference on Research and Development in Information Retrieval, Philadelphia, US, 296–303, 1997.

Henrich, A. & Robbert, G. (2000). *Combining Multimedia Retrieval and Text Retrieval to Search Structured Documents in Digital Libraries.* DELOS Workshop: Information Seeking, Searching and Querying in Digital Libraries 2000.

Kleinberg, J. (1999). Authoritative sources in a hyperlinked environment. *Journal of the ACM*, 46(5):604-632, November 1999.

Lynch, C. & Garcia-Molina, H. (1995). Interoperability, Scaling, and the digital libraries research agenda: A report on the May 18-19, 1995, IITA Digital Libraries Workshop, August 1995. http://www-diglib.stanford.edu/diglib/pub/reports/iita-dlw/main.html

McCray, A. T. & Gallagher, M. E. (2001). Principles for Digital Library Development. *Communications of The ACM*, May 2001, vol. 44, NO. 5.

Page, L., Brin, S., Motwani, R. & Winograd, T. (1998). The pagerank citation ranking: Bringing order to the web. Technical report, Computer Science Department, Stanford University.

Reddy, R. & Wladawsky-Berger, I. (2001). Digital Libraries: Universal Access to Human Knowledge – A report to the President. President's Information Technology Advisory Committee (PITAC), Panel on Digital Libraries. http://www.hpcc.gov/pubs/pitac/pitac-dl-9feb01.pdf, 2001.

Schatz, B. R. (1997). *Information Retrieval in Digital Libraries: Bringing Search to the Net*.

Sclaroff, S., La Cascia, M., Sethi, S. & Taycher, L. (1999). Unifying Textual and Visual Cues for Content-based Image Retrieval on the World Wide Web. *Computer Vision and Image Understanding*, 75(1-2):86-98, 1999.