

# Synchronised Object Retrieval: the enhancement of information retrieval performance in multimedia environments using synchronisation protocols

[Peter Brophy](#)

Centre for Research in Library and Information Management  
Department of Information and Communications  
Manchester Metropolitan University  
Manchester, UK

## Abstract

The retrieval of objects from within collections of multimedia presentations poses a number of problems but also offers opportunities for enhancing retrieval performance, by utilising information about the relationships between objects. This paper is concerned with the theoretical possibility of using the synchronisation information contained in SMIL-compliant multimedia presentations to retrieve objects which may either lack appropriate metadata or where the metadata is insufficient to enable reliable retrieval. It suggests that the synchronicity of display of objects could be used to infer their content and that this would provide possibilities for enhancement of retrieval performance. It further suggests how this process might be achieved and recommends that an experimental collection of SMIL-compliant presentations needs to be established to enable experimental work to be undertaken.

## Introduction

One of the roles of a research centre is to engage in speculative enquiry which may increase understanding of existing or future systems or open up possible new avenues of enquiry. This paper is concerned with one such exploratory study undertaken by CERLIM (the Centre for Research in Library and Information Management at the Manchester Metropolitan University) from 1999 to 2002. The study in question arose from observations of a variety of separate research and development projects in the digital library and information retrieval area (including CERLIM's [REVIEL](#) and [EDNER](#) projects), and concerned the question of how information objects expressed in a variety of media and perhaps embedded in multimedia presentations might be brought to the attention of end users. An obvious example would be an image or audio file which had been published as part of a complex multimedia presentation but which might contain information of relevance to a wide range of users who were not primarily interested in the complex object itself. An example might be a multimedia presentation on the Kalahari desert which incidentally contained an image of a sunset. How might an end user compiling a presentation on sunsets find such an embedded image?

While the example above may appear trivial, it illustrates what is likely to be a growing problem, namely that information objects will increasingly be 'packaged' in ways that make them invisible to standard methods of retrieval. Are there potential ways of addressing this problem in real world scenarios which do not simply involve the well-nigh impossible ? and probably unacceptable to rights-holders ? task of creating vast new datasets of disaggregated objects?

One approach, which has intriguing possibilities, would be to exploit the existence of documented relationships within complex objects. The most useful part of such documentation might be the information which is created to control the display of the complex object on the client workstation. In particular, it might be feasible to use information designed to enable the streamed delivery of multimedia across the Internet for the enhancement of

retrieval. The major problem with this type of delivery is that it cannot be taken for granted that each element of the presentation will arrive at the workstation when required. For example, because of network bandwidth or other limitations, it may be that a video track will be delayed while the accompanying audio arrives unhindered. Clearly, it would be unsatisfactory simply to play the audio and later, when it arrives, the video. To overcome this problem, the presentation includes a synchronisation track which not only tells the client the order in which to play individual components but contains check-points and other instructions on how to handle delayed or missing files.

One of the leading standards for synchronisation is called SMIL, the Synchronized Multimedia Integration Language, which is pronounced 'smile' ([World Wide Web Consortium, 2000](#)). This formed the environment for a feasibility study, carried out with funding from the UK's [Re:source, the Council for Museums, Archives and Libraries](#). This paper presents the findings of both that study (the Final Report is available online: [Brophy, et al., 2000](#)) and subsequent work to examine the ways SMIL is being implemented. Sample SMIL presentations and a tutorial are available online from [RealNetworks](#).

While for the purposes of the study reported here it was deemed useful to explore multimedia objects with explicit relationship coding, in principle there is no reason to suppose that the broad approach would not be more widely applicable. For example, it might be used to enhance retrieval of objects embedded in PowerPoint presentations by making an assumption that the individual slide itself signifies that a relationship exists between the different objects it contains.

## The concept of Synchronised Object Retrieval

Because SMIL defines temporal and other relationships between objects within a multimedia presentation, it should be possible to extract retrieval clues from the simultaneity of playback of constituent 'micro-objects'. The term 'micro-object' is used here to describe any mono-media file which is a component of a multimedia presentation. For example, if it is known that a text file is to be played back at the same time as an image file, and a text-based search of the text file retrieves relevant keywords, then if the same query expressed as a content-based search of the image file also produces possible matches, the inference can be drawn that

- the multimedia presentation as a whole has relevance to the query
- the instance identified (i.e. the simultaneous display of the text and image files) has relevance
- the text file has relevance
- the image file has relevance

and each of these conclusions may be drawn with a higher degree of confidence than a simple search of any one component part of the multimedia presentation could produce. In other words, the known synchrony of the files allows reinforcement of retrieval conclusions. We have termed this concept Synchronised Object Retrieval (SOR).

It is suggested that the SOR approach could be used either

- to enable the retrieval of individual micro-objects from within multimedia presentations across collections of heterogeneous multimedia presentations, or
- to enable the retrieval of multimedia presentations from one or more collection of such presentations, by examining the results of searching across some or all of a presentation's constituent micro-objects.

There are many possible applications of such capability ? an example would be that of a lecturer searching across a large collection of multimedia learning objects to find micro-objects to re-use in a new presentation.

In essence a SOR system would need to examine each multimedia presentation to determine its content, which might be expressed explicitly in metadata associated either with the complex object or with each micro-object or might be inferred by, for example, content-based image retrieval. The relationship between objects then needs to be examined to infer degrees of relevance.

As far as we have been able to determine the only related research in this area is that reported by Little, *et al.* where the objective is the dynamic generation of SMIL presentations by examining metadata and inferring **semantic** relationships. Temporal relationships in this work refer to, for example, dates of publication or of the life of a 'subject' or 'creator' of an object, not to temporal relationships in the display of the presentation:

Using the Open Archive Initiative (OAI) as a testbed, we have developed a service which uses the Dublin Core metadata published by the OAI data providers, to infer semantic relations between mixed-media objects distributed across the archives. Using predefined mapping rules, these semantic relationships are then mapped to spatial and temporal relationships between the objects.... Our premise is that by using automated computer processing of metadata to organize and combine semantically-related objects within multimedia presentations, the system may be able to generate new knowledge, not explicitly recorded, by inferring and exposing previously unrecognized connections. (Little, *et al.* [2002](#))

With large collections of multimedia presentations a SOR system would require a very high level of computing power. While in the past this may have been a reasonable objection to the SOR approach, the effects of [Moore's Law](#), which posits the inexorable doubling of computing power every eighteen to twenty-four months, suggests that computing power may not be an issue in the future; although recent commentary (e.g., [Tweney, 2002](#)) has questioned whether this trend will continue. However, while this may be the case in theoretical terms, it seems unlikely that computing costs for large-scale retrieval systems will not continue to fall or power per unit of expenditure increase for some time. Arms ([2000](#)) has made the observation that 'simple algorithms plus immense computing power often outperform human intelligence' and points out that processes which in the past would have been unthinkable are now routine, using the example of the Google Internet search engine:

Evaluating the importance of documents would appear to be a task that requires human understanding, but Google's ranking algorithm does remarkably well entirely automatically? Calculating the ranks requires the algorithm to iterate through a matrix that has as many rows and columns as there are pages on the web, yet with modern computing and considerable ingenuity, Google performs this calculation routinely.

## SMIL - the Synchronized Multimedia Integration Language

The Synchronized Multimedia Integration Language or SMIL standard is based on XML ([the eXtensible Markup Language](#)) and its syntax is defined in an XML DTD (document type definition). It is an official World Wide Web Consortium (W3C) standard. SMIL documents can be authored using a simple text editor, since they are in essence similar to HTML, although a variety of SMIL-enabled authoring tools are available. The project team investigated these and used the [RealSlideshow Plus](#) package as a basis for exploring the potential of SMIL. A demonstration SMIL application was developed during the study by CERLIM staff using this tool.

As part of the study the team undertook a detailed evaluation of mainly web-based resources relevant to SMIL. Because the concept being explored would require familiarity with recent research in text, audio, video and other media retrieval, a wide-ranging literature review was also undertaken and this has been reported elsewhere ([Hartley, et al. 2000](#)).

The SMIL standard defines a number of different types of media micro-objects that can be included in a presentation.

- animation: animated vector graphics or other animated format
- audio: audio clip
- img: still image
- text: text reference
- textstream: streaming text
- video: video clip
- ref: generic media reference (in effect, 'other')

For example, an author can write synchronisation instructions such as: 'play audio file A in parallel with video file B' and 'display text file C after audio file A and in parallel with animation file D and with text stream E'. In addition SMIL enables the author to define the positioning of micro-objects on the user's screen (or 'visual rendering surface' in the terminology of the standard). Various other variables can be used, such as the ability to test the end-user's bandwidth and deliver alternative micro-objects accordingly. However, these features are unlikely to be of major significance in the context of an information retrieval system.

One of the advantages of the SMIL approach is that it effectively enables objects which would normally require

high-bandwidth, such as multimedia presentations, to be delivered across low-bandwidth networks since each constituent micro-object can be transmitted separately and the presentation correctly reassembled and played at the client end, using the synchronisation information. It thus has enormous potential for the distribution of all kinds of multimedia objects. Furthermore, there is no reason why the micro-objects cannot be repackaged and delivered within separate presentations, without the expense of re-recording a video or other complex element as might be the case were the presentation to be recorded as a single file of data. Among other useful aspects of the approach is the use of alternative files: for example, a commentary can be stored in a variety of languages and the user given the option as to which should be played, or the option can be picked up automatically from the user's preference settings. SMIL 2.0 (the current standard) has introduced the concepts of modularisation and profiling to enable functionality to be extended and the integration of related mark-up languages.

## SMIL and information retrieval

While SMIL has not been designed with information retrieval (IR) applications in mind, the schema does contain some explicit IR-related features, most notably the provision of metadata fields at both object and micro-object levels. Thus, the standard defines a [Resource Description Framework](#) (RDF) compliant 'metainformation' module which should enable the macro-object and its constituent micro-objects to be retrieved in the same way as any other documents by examining metadata elements. While the schema used is open, the SMIL 2.0 standard suggest the use of [Dublin Core](#) (DC) as a generalised approach and an example of this usage can be found in section 8.5 of the standard. In addition to, say, subject-based retrieval, using the DC subject, title and description elements for instance, a SOR system could use some elements to refine search results. Using the DC 'creator' element to find micro-objects which were created by the same person so as to infer some commonality might be an example of this approach. There is also a field intended for content rating metadata, such as the [Platform for Internet Content Selection \(PICS\)](#), which, in some circumstances, might be a useful aid to retrieval.

However a difficulty, as our later investigations showed, is that subject metadata fields are frequently unused by multimedia producers, especially at the micro-object level, or where they are used there is no meaningful vocabulary control. Thus while these fields would be examined in any real-life SOR application, it is to other features of SMIL that developers would need to look. Among those relevant to IR are:

- SMIL supports the inclusion of linked objects in a presentation, through its Linking Module, enabling events such as a user clicking on an area of the screen, to trigger a link to another part of the presentation. As in HTML, provision is made for supporting name fragment identifiers, by using the # symbol, to enable a link to be made to any tagged element within a micro-object. For example, if the SMIL 2.0 documentation was itself part of a SMIL presentation, a link could be provided to <http://www.w3.org/TR/smil20/extended-media-object.html#edef-ref> where definitions of media object elements would be found. If this is not the start of the micro-object the effect would be as if the user had fast-forwarded within that object to a specified place. Further refinement is possible: for example, SMIL contains provision to enable the author to specify precedence levels for cases where the origin and linked objects continue to play simultaneously - an example would be setting the relative audio playback levels. Again, although we did not explore these in detail in the feasibility study, it would be possible for software to follow embedded links and to take account of associated variables to retrieve further micro-objects which may themselves contain both usable metadata and other retrieval clues.
- The SMIL 2.0 standard has specified various features intended to improve accessibility, especially for people with visual or other disabilities. For example a 'long description' attribute enables the author to provide a long text description, separate from the metadata fields, which could be very useful for enhancing retrieval. Because of the ability to define user-required variants it is also possible to allow for the identification of, for example, text equivalents of audio intended for closed-captioning, thus providing a further accessibility feature and additional IR possibilities.
- As noted above, a 'language' attribute is available, intended for presentations where the user has choice of, say, a French, German or English sound track: the multimedia presentation, of course, would contain three files, each synchronised with the video and other elements. Again, the possibilities for enriching IR capability, by combining searches in a variety of languages, are intriguing.

Finally, it should be noted that some experimental work has been carried out on the building of SMIL compliant multimedia archives from the perspective of metadata **creation** ([Hunter and Little, 2001](#)) although this does not address the use of relationships between micro-objects for retrieval.

# Experimental Design

Although the CERLIM team has not yet developed the SOR concept beyond consideration of feasibility, initial work has been carried out to define the requirements for moving the SOR concept into an experimental stage. We believe that the elements described in the following sub-sections would be required.

## Establishment of test collection

It will be necessary to establish a reasonably-sized test collection of SMIL-compliant multimedia presentations, showing a degree of heterogeneity in structure and micro-object composition so as to challenge applications. Despite examining a number of applications we have not been able to identify a suitable collection of SMIL-compliant objects at the present time, although the work of Little, *et al.* referred to above must be noted. We also note that some candidate collections are restricted by IPR considerations.

It will be necessary to establish a test collection along the lines of [TREC](#), the Text Retrieval Conference. In order to test the SOR approach thoroughly in experimental conditions, a carefully designed collection would be needed. In other words, it would be essential to put effort into ensuring that the collection displayed examples of all the different elements on which retrieval could be based, but with special emphasis on synchronicity. Thus, not only would examples of objects with comprehensive metadata at the object and micro-object levels be needed, but care would need to be taken to ensure that the different media used contained retrievable content, the attributes of which were known, in order that testing of software and algorithms could take place.

Because of the complexities inherent in the SOR approach, our conclusion is that a test collection should be domain-limited initially. For example, the use of a domain such as, say, the music of a specific composer or a specific locality's history and culture would avoid the worst trans-domain semantic and other problems. The examples that Little, *et al.* ([2002](#)) provide in their paper also show this concentration on a specific domain.

The development of a test collection would be a challenging piece of work in its own right, but would be an essential first step to the establishment of an experimental programme. It would be helpful to consult with the TREC team in defining this collection and it may be that the [TREC video track](#) would form a good starting point.

It is perhaps worth noting here that although the test collection would need to be carefully crafted, this does not mean that it would have to be entirely artificial. There would be merit in identifying a nascent, real-life collection which could be SMIL-enabled, not least because experimental work could then contribute to the solution of real-world problems and real-world queries could inform the evaluation of retrieval.

## User queries

The analysis of the user query into mono-media components will itself be a complex process. Although in theory this could be automated, so that software converted a query expressed, let us say, as a text string ('Find something about the sun') into a series of media specific queries ('sun' [text]; 'sun' [audio wave file, retrieved from a 'thesaurus']; 'sun' [image description, 'round yellow object on blue background', retrieved from another 'thesaurus']; and so on), in practice this is currently impractical or at least a different area for research. In practice it may be feasible to generate a multi-media query by providing the user with an interface which enabled descriptions to be selected/entered in a series of 'channels' e.g. text box, select from image thumbnails to find 'things like', audio input terms and so on, although the problem of segmenting and recognising audio-based queries is non-trivial, and non-verbal audio, the sound of waves breaking on a beach, for example, poses even greater problems than speech.


It would be possible to make user input an iterative process by presenting to the user examples of retrieval terms, especially where non-textual retrieval was being undertaken. Thus, to extend the above example, an initial user query for an image of the sun might result in a series of thumbnails, extracted by reference to a thesaurus, being displayed and the user being invited to select from them or rank them in order of relevance. This additional user input could also be used to enable the system to learn from search preferences and possibly provide personalisation of results, although again this would be beyond the scope of the initial experiment.


As part of the feasibility study we developed a mock-up of the type of user interface we have in mind. This is





shown in Fig. 1.


---



☐ Select


☐ Select

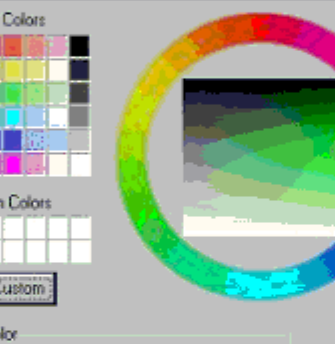

☐ Select


☐ Select


☐ Select


☐ Select

---



---

☐ Sound

☐ Motion

---

Select any combination of the above to create your search

**Figure 1: Mock-up of a user interface**

## Query Analysis

The user query is analysed into query statements capable of being applied within specific mono-media contexts as discussed above. So, for example, the query might be expressed as a series of text strings, some of which might be derived from speech, for matching against a text file, as a voice/sound wave format for matching against an audio file and so on. A later project might examine how, for example, text strings might be applied to an intermediate image thesaurus to create image queries not present in the original input data. As indicated above a possible way to use such query enhancement would be to display a set of images to the user and invite selection of the nearest matches. Again, we would expect this phase of work to draw on and be influenced by a range of research, especially

in the content-based information retrieval (CBIR) field.

## **Search of presentations**

The first process should consist of a search of the metadata associated with each multimedia presentation as a whole. A ranking of objects in the collection could be achieved in this way, although it would be important not to discard presentations with apparent zero relevance from this process, since many objects may lack metadata and much presentation-level metadata may be irrelevant to specific component micro-objects. An example of this could be a presentation designed to illustrate a geographical area which contained images of flora and fauna of the region but described merely by geographical location ? the metadata would probably produce a 'zero relevance' result for a query for a particular plant or animal even if they were in fact represented in micro-objects.

This process leads to a score 'A' being computed for each presentation. As with other searches, this stage is intended to reveal clues, not to stand on its own as a retrieval process. Where some presentations do not have associated metadata it would be appropriate to provide a neutral score.

## **Search of micro-objects' metadata**

The search of the micro-objects themselves could be carried out in a number of ways. For simplicity we suggest that the first step should be to check all micro-objects for embedded metadata, to search on this and to rank micro-objects for relevance accordingly. Again, many micro-objects will lack metadata so again all that is produced is a set of clues to relevance. This process allows a score  $B_x$  to be computed, where  $x$  is the micro-object sequential number. It should be noted that micro-objects may be regarded as being in any order for this purpose, provided only that they can be processed sequentially. Thus it may be appropriate to treat them as sequenced by filename, by file type, by 'running order' or in a variety of other ways. To avoid micro-objects without metadata being given the lowest scores it would again be appropriate to calculate a mean score for all micro-objects which have metadata and use this to ensure that metadata-less micro-objects score neutrally. It would of course be necessary to identify the most appropriate scoring model for the purpose.

## **Search of micro-objects' content**

Here we reach the core of the approach. We assume here that the aim is to retrieve micro-objects, although the process of retrieving presentations is similar, requiring only a method for combining the retrieval scores of constituent micro-objects. The process might take the following base model:

1. The synchronisation file is analysed to identify the first micro-object ? note again that 'first' here is arbitrarily defined, simply implying that there is some ordering which ensures that each micro-object will be examined in sequence. The media type should be defined in the SMIL type attribute and this should be used to determine the type of retrieval process to be engaged. Using this process, micro-object 1 is examined, using a retrieval method other than examining metadata, as this has already been undertaken, and a score  $C_1$  computed. Again, the establishment of a scoring algorithm would be an important part of the experimental research.
2. Micro-objects which are due to be played simultaneously with the first object (this should be interpreted as meaning that there is some temporal overlap, not that they start and end at the same time) are then examined against whatever media retrieval process is appropriate. The results of each analysis are used to modify  $C_1$ , producing  $D_1$ : again we have not defined the algorithm to be used, although it may be noted that different media types might be weighted differently if it was found that, say, text retrieval produced more reliable results than content-based image retrieval. The effect of this stage is that  $C_1$  is increased whenever a high probability of a match in a micro-object due for simultaneous play is identified.
3. When micro-object 1 and all other micro-objects with which it shares some temporal space have been processed, the next micro-object in the sequence is examined and steps 1 and 2 repeated, providing scores  $D_2$ ,  $D_3$ ,  $D_4$  ???..  $D_n$ . It should be noted that because the synchrony is defined for each micro-object, it will be necessary to use the relevance score for each multiple times ? clearly it will be more efficient to store these and re-use them during the process than to recompute them.
4. At the end of the object analysis a relevance score for each micro-object is available. This is combined - again, experimentation will be needed to establish the algorithm to be used - with the overall score  $A$  and the metadata score  $B_n$  to produce a 'final' score for each micro-object,  $E_n$ .

5. When all objects in the collection have been analysed, the micro-object scores  $E_n$  can be ranked to indicate the most likely matches overall. These micro-objects can then be displayed to the user.

Where the object is to retrieve multimedia presentations rather than micro-objects, clearly a score for each macro-object can be computed by combining - again with a suitable, yet to be defined, algorithm - the scores for all constituent micro-objects and then ranking the overall score against those for other presentations.

## Conclusions

The examination of the SMIL standard and its application described in this paper demonstrates that even in the absence of explicit metadata, or where the quality of available metadata is poor, it should be possible to develop systems which use information on the relationships between micro-objects to achieve enhanced retrieval performance. The standard also contains features, such as those intended to improve accessibility, which should be explored from a retrieval perspective.

It is suggested that the next stage of research should involve the development of a suitable test collection of SMIL-compliant material and the establishment of an experimental programme to enable the ideas explored in outline in the feasibility study to be subjected to in-depth research and development. It is the intention of CERLIM to pursue funding to enable these programmes of research to be established as soon as possible.

It is clear from our research that SOR has enormous potential for application in networked information environments containing heterogeneous collections of multimedia presentations. As such environments proliferate, as the number of multimedia presentations mushrooms and as the cost of computing power reduces, it is to be expected that SOR approaches will prove ever more valuable.

## Acknowledgements

The author wishes to acknowledge the contributions of many colleagues in CERLIM and the Department of Information and Communications to this work. Particular thanks are due to the late Tony Oulton, who undertook much of the background literature review, and to Richard Eskins who set up SMIL demonstrators and designed the mock-up of a query entry page which is illustrated here in Figure 1.

## References

- Arms, W.Y. (2000) [Automated digital libraries: how effectively can computers be used for the skilled tasks of professional librarianship?](http://webdoc.gwdg.de/edo/d-lib/dlib/july00/arms/07arms.html) *D-Lib Magazine*, 6(7/8) Retrieved 10 February 2003 from <http://webdoc.gwdg.de/edo/d-lib/dlib/july00/arms/07arms.html>
- Brophy, P., Eskins, R. and Oulton, T. (2000) [Synchronised Object Retrieval: a feasibility study into enhanced information retrieval in multimedia environments using synchronisation protocols](http://www.cerlim.ac.uk/projects/synchro/finalv2.pdf) Manchester: Centre for Research in Library & Information Management. (Library and Information Commission Research Report 92) Retrieved 17 July 2003 from <http://www.cerlim.ac.uk/projects/synchro/finalv2.pdf>
- Hartley, R.J., Johnson, F.J. and Oulton, A.J. (2000) Image, audio, text: a review of recent research in information retrieval. *The New Review of Information and Library Research*. 6, 171-206.
- Hunter, J. and Little, S. (2001) [Building and indexing a distributed multimedia presentation archive using SMIL](http://archive.dstc.edu.au/RDU/staff/jane-hunter/ECDL01/ECDL01.html). In: P. Constantopoulos and I.T. Solvberg eds. *Research and advanced technology for digital libraries. 5th European Conference, ECDL 2001, Darmstadt, Germany, September 4-9, 2001. Proceedings.* (pp. 415-428) Heidelberg: Springer Verlag. Retrieved 17 July 2003 from <http://archive.dstc.edu.au/RDU/staff/jane-hunter/ECDL01/ECDL01.html>
- Little, S., Guerts, J. and Hunter, J. (2002) [Dynamic generation of intelligent multimedia presentations through semantic inferencing](http://archive.dstc.edu.au/maenad/ecdl2002/ecdl2002.html). In: M. Agosti and C. Thanos, eds. *Research and advanced technology for digital libraries. 6th European Conference, ECDL 2002, Rome, Italy, September 16-18, 2002. Proceedings.* (pp. 158-189). Heidelberg: Springer Verlag. Retrieved 26 May 2003 from <http://archive.dstc.edu.au/maenad/ecdl2002/ecdl2002.html>
- Tweney, D. (2002) [Does Moore's Law still hold true?](http://dylan.tweney.com/writing.php?display=331) [Originally published in *Business 2.0*] Retrieved 17 July 2003 from <http://dylan.tweney.com/writing.php?display=331>
- World Wide Web Consortium (2000) [Synchronized Multimedia Integration Language \(SMIL 2.0\)](http://www.w3.org/TR/2000/REC-smil20-20001214/)



**Find other papers on this subject.**

---

**How to cite this paper:**

Brophy, P. (2003) "Synchronised Object Retrieval: the enhancement of information retrieval performance in multimedia environments using synchronisation protocols." *Information Research*, **8**(4) paper 161 [Available at <http://InformationR.net/ir/8-4/paper161.html>]

---

? the author, 2003.  
Last updated: 26 May 2003



---

Check for citations, [using Google Scholar](#)

---

[Contents](#)

**3 1 8 4**  
[Web Counter](#)

[Home](#)

---