Using Ontologies to Retrieve Video Information: A System Proposal

Christian Paz-Trillo[1], Renata Wassermann[1] and Paula P. Braga[2]

[1] Department of Computer Science
Institute of Mathematics and Statistics
University of São Paulo, Brazil
[2] Department of Philosophy
University of São Paulo, Brazil

Abstract. Searching for information in long videos can be a time-consuming experience. This paper describes a project oriented towards efficient information retrieval from digital video. The prototype being implemented uses an interview with a Brazilian artist. The system allows the user to enter questions to the artist in natural language (Brazilian Portuguese) and selects the pieces of video that best answer the questions. The retrieval is based on an art ontology which is currently being developed. The system is intended to be used in museums and art galleries in order to provide the visitor with a more interactive interface than the usual wall or catalogue texts.

1 Introduction

In this paper, we describe an application of ontology-based video retrieval for museum and art exhibition spaces. We recorded interviews with Ana Teixeira, a Brazilian artist whose work deals mainly with urban interventions. Our project intends to implement video retrieval through queries in natural language (Brazilian Portuguese) and allows for a new relationship between the art spectator and the work of art: the visitor of a museum or art gallery will be able to develop his/her own questions and thoughts about the exhibited works. Exhibition spaces usually offer only wall or catalogue texts to the visitors, frequently written in a very specialized language, and focusing on the curatorial thesis, which leaves no room for the spectator's creative thought. In some cases, such as contemporary art works that privilege the participation of the spectator on the making of the work, such museological practices represent an unacceptable gap and an inconvenient mediation between the goals of the artist and the experience of the museum visitor. Ana Teixeira develops her work on the streets or other public spaces, frequently involving passers by in the work. Bringing her work to a conventional exhibition space asks for new approaches to the relationship between the visitor and the museum. Our system establishes a dialogue between the spectator and the artist, which is very desirable in the context of contemporary art.

The interviews with Ana Teixeira were recorded in digital video and served as a guide to the development of the first version of our contemporary art ontology, an ongoing process that will continue as we record interviews with other

contemporary artists. We are interested in obtaining a flexible system, that can be configured to other interviews, or to another kind of videos (courses, seminars, etc.) by extending or changing underlying ontologies and organizing a clips database for it.

Section 2 introduces the Information Retrieval problem and approaches to solve it that will be used in this project. Ontologies that are being developed are described in Section 3 as well as the tools being used to create and access them from the system. In Section 4 we describe the modules to be implemented and how they relate to each other, and show what is currently implemented. In Section 5 we present related works on video retrieval and ontology-based information retrieval. Section 6 describes the results we expect to get from the implementation in terms of what is going to be measured. Finally, in Section 7, we present our concluding thoughts and what remains to be done after the proposed implementation is completed.

2 Information Retrieval

2.1 Definition

An Information Retrieval (IR) System allows users to look for information in a database through queries usually formatted as a set of keywords (or indexing terms)[BYRN99]. Using the query, an IR system retrieves information that might be relevant to the user.

2.2 IR problems

Keyword-based IR systems are limited in its ability to distinguish between relevant and irrelevant texts [Mau91], mainly due to:

- **Synonymy:** This is the case when there are several terms to describe the same object (or concept). A keyword-based IR system will only retrieve those documents that refer to the object (or concept) by the same term used in the query. For instance, a query to look for information about accommodations in Rio de Janeiro could be "Inn Rio de Janeiro", and only documents containing the word "Inn" would be retrieved.The system would miss information associated with words such as "hotel" or "hostel".
- **Polysemy:** A polysemic word is a word with multiple meanings, i.e., expressing different concepts. In this case a query might lead to the retrieval of documents which deal with concepts foreign to the subject of interest. A query for "tree" in the sense of data structure might retrieve documents related to "tree" in the sense of a biological organism.

We refer to these limitations as the *keyword barrier* [Mau91]. To go beyond this obstacle, an IR system needs another structure for the data with which it works, e.g., an structure based on concepts instead of mere keywords. The concepts are

organized in a relational and hierarchical structure, an ontology, that solves the synonymy problem. We will not deal directly with polysemy, but the domain-oriented vision of the system diminishes its relevance and impact on the system.

2.3 Performance Measures for IR systems

There are standard measures to evaluate the performance of IR systems [Mau91]:

- **Precision:** The ratio of documents retrieved by the system that are actually relevant to the query divided by the total number of documents retrieved.

$$P = \frac{\text{Number of relevant documents retrieved}}{\text{Total number of documents retrieved}} \tag{1}$$

  For instance, if the system retrieved 6 documents for a query, where 3 of them were actually relevant, the precision performance for the system in that query is 0.5 or 50%. Polysemy may produce low precision rates, because irrelevant documents might be retrieved.
- **Recall:** There may be many documents in the database that the user considers relevant, but only some of them will be retrieved by the system. The recall performance of a query is the ratio of the number of relevant documents retrieved by the system divided by the total number of relevant documents in the database.

$$R = \frac{\text{Number of relevant documents retrieved}}{\text{Total number of relevant documents in the database}} \tag{2}$$

  Recall is difficult to measure, because it requires knowledge about the total number of relevant documents in the database, which has to be determined manually. To measure recall, an upper bound is usually established by finding documents that should have been retrieved but were not. The greater the manual effort to find these documents, the more precise the recall measure is. Synonymy leads to lower recall rates, because relevant documents referring only to synonyms of a word used in the query may not be actually retrieved.
- **Response time:** The elapsed time between the submission of a query and the presentation of the documents retrieved by the system.

Precision could be easily maximized by retrieving a single document that is certainly relevant, and recall by retrieving all documents in the database. Thus, a measure that combines both of them is preferred, for example, the*F-measure* [vR79]

$$F = 2\frac{R\ P}{R + P} \tag{3}$$

where F-measure is the harmonic mean of precision and recall. The advantage of using F-measure is that maximizing it means maximizing a combination of recall and precision.

2.4 Using Ontologies in IR

Ontologies have been commonly used in Information Retrieval to improve recall and precision. Two approaches have been used to achieve these improvements: query expansion through the use of semantically related concepts, and conceptual distance measures.

The first approach expands the user query by adding terms semantically related to those used in the original user's query, so documents that do not necessarily contain terms present in the query may be retrieved. The second approach uses a conceptual distance to measure the similarity between terms in a query and terms in a document. An ontology can be seen as a directed graph, where the nodes are the concepts in the ontology, and a directed edge between two nodes is drawn where exists any relationship between those two concepts in the ontology. Therefore, the conceptual distance between two concepts is the length of the shortest path between them in the directed graph.

3 Domain Ontology

We are developing an ontology that contains the concepts used in the interview with Ana Teixeira. The concepts refer mainly to people, institutions and objects related to contemporary Brazilian art. We also included concepts related to more general ideas like feelings, countries or places, where ontology reuse could have helped our work. However, because the system works with Brazilian Portuguese and most shared ontologies are developed in English, little ontology reuse was possible. In the future, we intend to turn it into a bilingual ontology supporting both English and Brazilian Portuguese.

3.1 Development Process

The exercise of developing an ontology about contemporary art raised several philosophical thoughts we consider important to mention here. Today art deals with the effacing of frontiers regarding classifications. It is impossible to constrict a work of art to the domain of a category like "painting" or "sculpture". New classifications appear from time to time to work around this language limitation: one may classify a work of art as "performance" because the artist paints a canvas while dancing around it during the exhibition, or "installation" because the spectator is confronted with an environment that includes music, video, sculptures, smells, all making a single work of art. The unavoidable overlapping of categories and constant linguistic efforts to isolate artistic manifestations and alternative thought matrices into well-defined and comfortable cubicles are the heritage of the Classic notion of reason, that supposes we need a method to conduct us to true knowledge. Contemporary philosophy and art have, however, disrupted the notion of truth and knowledge. Can we really know the world or are we limited to what we can apprehend with our human linguistic tools?

Modeling the art world through artificial intelligence techniques such as ontologies and IR systems is a thought-provoking exercise. By dealing with this

paradox - awareness of the limitations of the Classic category-based system of thought versus the necessity of developing a machine-oriented, concept-based model of the art world - we highlighted the urge to efface categories in human thought: this project is both an artificial intelligence system and a philosophical exercise.

Limiting the world to a machine-understandable system of concepts and categories, we started modeling the art world with jargon terms such as "institution", "work of art", "place", "text", "individual", and dismembering each of these main blocks into sub-classes in a "type of" relationship. Thus, an "institution" could be a museum, a gallery, a cultural center, a company, a school, and so on. Later, we examined all the key-words we extracted from the interviews with the artist Ana Teixeira, detecting the necessity to include additional super-classes such as "Material" and "Power". We tried to have classes to encompass all the words she mentioned in her talk that the domain expert considered relevant. The next step was to associate a list of synonyms to each word that appeared in the class tree[3]. Later, we developed relationships between classes to accommodate terms that the class system missed and we are currently registering instances into our model.

This procedure will be repeated as we interview other contemporary artists, detailing more and more the ontology. The flexibility of the ontology system allows us to reevaluate our model from time to time, a necessary step in any system of thought. For details about the current state of our ontology, see `http://www.ime.usp.br/~liamf/BIBO/ontologies/arte/`.

3.2 Ontology Tools

We are using `Protégé-2000` [GMF+03, Pro04] to elaborate the ontology, because its `OWL`-plugin [KMR04] allows exporting ontologies into the Web Ontology Language[SWM04](`OWL`). `OWL` is the W3C recommendation for describing ontologies, derived from the Resource Definition Framework[Con04](`RDF`). A screen shot of the current state of the ontology inside the `Protégé` environment is shown in Fig. 1, it shows part of the hierarchy below the class "Obra de Arte" (Work of Art).

To explore the ontology from inside the system, we are using `Jena`[Lab04], a Java Framework that has a comprehensive subsystem to manipulate ontologies, and that allows for `OWL` manipulation.

4 System Description

Figure 2 shows the different modules that are being implemented for the system.

[3] Synonyms are represented as commentaries of the class, to facilitate maintenance, but after an automatic processing they will be converted to a language construction that establishes an equivalence relationship between two classes.

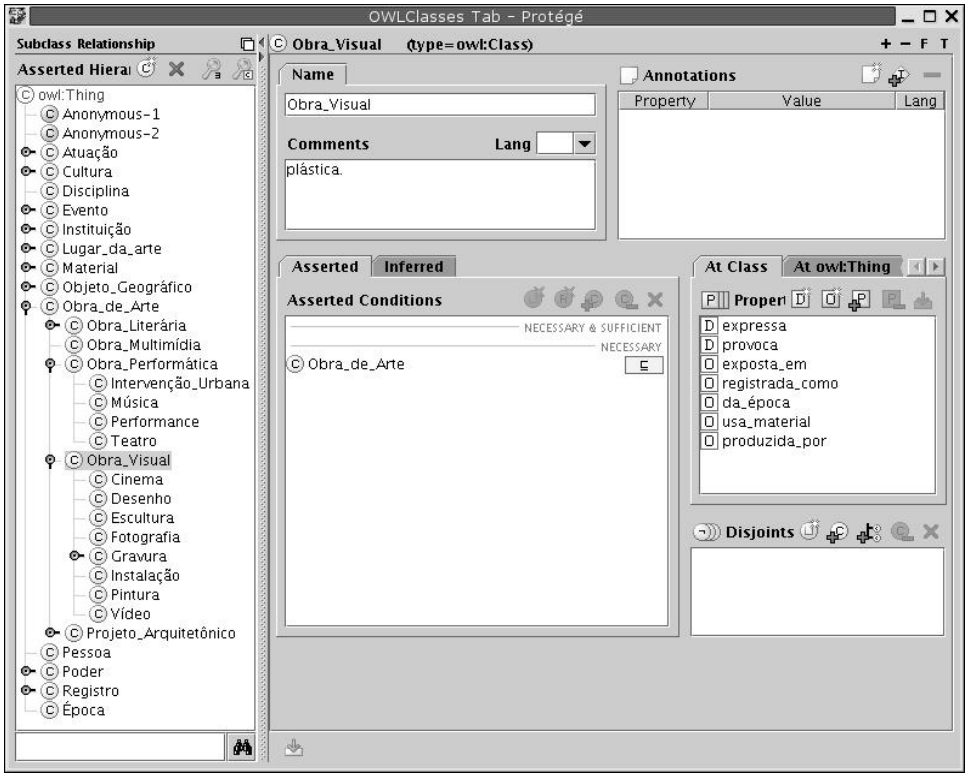

Fig. 1. Ontology screen shot.

- **Preprocessor:** The natural language query will be processed in order to detect and correct misspelled words. This misspelling correction will be made using `Jazzy` [IR04], an open source spell checker, using an electronic dictionary of Brazilian Portuguese, `br.ispell` [Ued02], and a domain-oriented dictionary automatically extracted from the ontology in the format supported by `Jazzy`.
  The preprocessor will then apply a stemming process to avoid overloading the retrieval process. Stemming is the process of removing affixes, i.e., prefixes and suffixes, of the word. A stem is a word after this removal. A generic and widely used stemming algorithm was proposed by Porter in [Por80], and it was refined to be used with Brazilian Portuguese in a work described in [OH01]. We will use this refined algorithm in the preprocessor.

- **Morpho Syntactic Tagger:** The preprocessed query is passed to this component that identifies the syntactic components of the query relevant to the video retrieval. Question words like "quem" (who), "qual" (which) or "como" (how) will be tagged, so the objective of the question will be known, e.g., "quem" questions may refer to people. Words that are present in the domain-

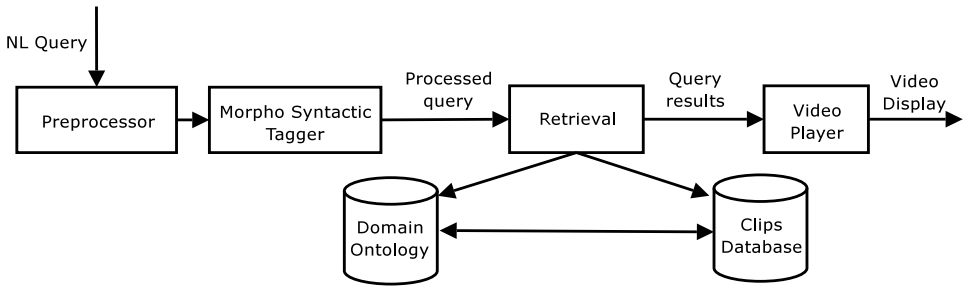

Fig. 2. General architecture of the system.

oriented dictionary will also be tagged.

- **Domain Ontology:** The domain ontology being developed was explained in Section 3.

- **Clips Database:** The clips are stored in MPEG format. An `XML` file contains information about clips and will also determine the relationships between each clip and the concepts it references. Additionally it relates clips to resources (photographs or figures) that are shown at a specified time within the video player, usually while interviewee talks about them. This database will be used in the retrieval process to get the clips relevant to the query.

- **Retrieval Process:** The retrieval will take the tagged query and look for concepts semantically related to the words in the ontology. The clips that are nearest to the query concepts will be searched in the clips database, ranked, and displayed to the user. Related clips will be ranked according to its conceptual distance, and the kind of query detected by the Morpho Syntactic Tagger, e.g., looking for people or places.

- **Video Player:** The application will show the ranking of retrieved clips and will allow the user to watch any of them. The video player is already implemented and uses the *Java Media Framework*[Sys04](`JMF`) to exhibit video clips and allow the user to control them through play, pause, and stop buttons and to browse the results through next and previous buttons.

The clips database as well as the retrieval process and the video player are already implemented to support keyword-based video retrieval as a `Java` application. This application allows the user to retrieve the clips based on keywords, but, as explained in Section 2, this approach does not deal with synonymy, and specifically this application does not make any preprocessing, so making a query by "artistas" (artists) could retrieve different results than a query by "artista" (artist) for example. A screen shot of the application is shown in Fig. 3.

Fig. 3. Query by "artista" in the current implementation.

The current keyword-based implementation will be compared to the proposed concept-based system in order to assess the benefits of using ontologies.

5 Related Works

As explained in Section 2, ontologies have been commonly used in Information Retrieval. OntoSeek [GMV99] proposed the use of ontologies to increase both recall and precision in narrow domains, such as product catalogs. OntoSeek used a limited language to represent concepts and a large ontology, WordNet [Mil95], for concept matching. Khan proposed, in [Kha00], the ideas of conceptual distance were applied to retrieve audio data using a query expansion mechanism that deals with natural language user queries in the domain of sports. In terms of techniques used for retrieval this is the work most related to ours.

Some work related to automated content-based video retrieval was developed at Cambridge University, and published in [BFJ+95], which presents a statistical approach for browsing multimedia documents, specifically broadcast news video. Other works use MPEG-7[4] to provide more detailed information about the content of audio or video data, and exploit it to make the retrieval[BMN+04].

An use of ontologies for video retrieval was proposed by the ISIS research group from the University of Amsterdam [WBvG+02], offering an interactive

[4] MPEG-7 is a standard for description and search of audio and visual content.

method for video retrieval by guiding the user's interaction with domain information extracted from an ontology.

In January, 2001, The Krannert Art Museum in Champaign, IL, USA, featured a large exhibition on the work of sculptor Jacques Lipchitz (1891-1973). Besides sculptures, drawings, and paintings the show presented a software developed by Bruce Bassett and Histor Systems entitled "Conversartion with Jacques Lipchitz: A Breakthrough in Interactivity". This system started to be developed in the 1970s and, in its first version, worked with VHS tapes that were manually selected and played according to the asked question. The version presented at the Krannert Art Museum in 2001 upgraded this pioneering work by using digitized video fragments. However, since it was part of a commercial enterprise we have little information about its implementation. It might have used ontologies. In any case, Bassett's work with the cubist sculptor, which summed up to 300 hours of taped interviews, is certainly a precursor to what we developed. It nicely broke with traditional exhibition practices and provided a very attractive interaction between the artwork and the spectators, bringing back the thoughts on art of the deceased sculptor[Kli01, BS01].

6 Expected Results

The system in its current state, i.e., a keyword-based video retrieval system will be compared to the ontology-based system currently under development. The tests will consist of a set of queries, both by keywords and natural language queries. Precision, recall and response time will be measured and we will compare the F-measure for both implementations. We expect that the concept-based video retrieval will outperform the keyword-based system.

We also expect our system to be actually installed in museum and gallery spaces, as a substitute for wall texts, and thus to provide a less authoritarian mediation between the spectator and the ideas of the artist.

7 Conclusion and Future Work

Ontologies will help to deal with synonymy thanks to the query expansion mechanism that will be used. The construction of an ontology is an iterative process. Specifically in our case, the ontology is mainly related to the topics of the interview, which involves just a subset of the domain of Brazilian contemporary art. Then, it will become a first version of an ontology that will be more and more detailed, as more interviews are done.

Although this project is an application of ontologies to art exhibition environments, it intends to offer a framework easily adapted to other subjects. In order to do that, all one needs to do is to develop another ontology and video database. For example, a whole course on any subject can be recorded in video and then accessed through specific questions proposed by the user, who would see, in addition to the lecture audio and video, figures, graphs, and slides, in the

same way we currently alternate images of our interview with photographs of related works of art.

Having the interviews recorded in digital video, a domain expert needs to fragment the video in smaller clips and associate concepts to each clip, a time consuming process that could be done automatically. In doing that, the system would depend on the domain expert only for the development of the ontology. To fragment the video automatically, it would be necessary to do a preprocessing to identify all the words mentioned in the video and, comparing them to the ontology concepts, to fragment the video in smaller clips. Speech recognition could also be added to the system to receive the user's input question.

The system would also benefit from an administrative interface that would help the domain expert to add or remove concepts associated to each video fragment.

Acknowledgements: Paula P. Braga is supported by a grant from CNPq, Process # 140923/03-9. Christian Paz-Trillo is supported by CAPES. Renata Wassermann is partially supported by the CNPq grant PQ 300196/01-6. This work has been supported by CNPq project 55.0222/2003-0.

We would like to thank Fábio Kon for his technical review and ideas given to this project and to the anonymous referees for the comments.

References

[BFJ+95] M. G. Brown, J. T. Foote, Gareth J. F. Jones, K. Sparck-Jones, and S. J. Young. Automatic content-based retrieval of broadcast news. In *Proceedings of the Third ACM Multimedia Conference*, pages 35–43, 1995.

[BMN+04] W. Bailer, H. Mayer, H. Neuschmied, W. Haas, M. Lux, and W. Klieber. Content-based video retrieval and summarization using MPEG-7. In *Proceedings of the Internet Imaging V*, pages 1–12, San Jose, CA, USA, January 2004.

[BS01] B. Bassett and Histor Systems. Conversation with Jacques Lipchitz: A breakthrough in interactivity, 2001. A description of the system is available in: http://www.conversationwithjacqueslipchitz.org/.

[BYRN99] R. Baeza-Yates and B. Ribeiro-Neto. *Modern Information Retrieval.* Addison Wesley Longman, 1999.

[Con04] World Wide Web Consortium. RDF Resource Definition Framework, 2004. http://www.w3.org/RDF/.

[GMF+03] J. Gennari, M. Musen, R. Fergerson, W. Grosso, M. Crubézy, H. Eriksson, N. Noy, and S.Tu. The evolution of Protégé–2000: An environment for knowledge–based systems development. *International Journal of Human-Computer Studies*, 58(1):89–123, 2003.

[GMV99] N. Guarino, C. Masolo, and G. Vetere. Ontoseek: Content-based access to the web. *IEEE Intelligent Systems*, 14(3):70–80, May 1999.

[IR04] M. Idzelis and A. Roy. Jazzy - Java Spell Check API, 2004. http://sourceforge.net/projects/jazzy.

[Kha00] L. Khan. *Ontology-based Information Selection.* PhD thesis, Department of Computer Science, University of Southern California, 2000.

[Kli01] G. Kline. High-tech sculptor has the answers. *The News-Gazette Online*, October 2001. Published online October 4, 2001, available in: http://www.news-gazette.com/story.cfm?Number=10249.

[KMR04] H. Knublauch, M. Musen, and A. Rector. Editing description logics ontologies with the Protégé OWL plugin. In *International Workshop on Description Logics*, Whistler, BC, Canada, 2004.

[Lab04] HP Labs. Jena 2.1 - A Semantic Web Framework, 2004. http://www.hpl.hp.com/news/2004/jan-mar/jena2.1.html.

[Mau91] M. Mauldin. *Conceptual Information Retrieval A case study in adaptive partial parsing.* Kluwer Academic Publishers, 1991.

[Mil95] G. Miller. Wordnet: a lexical database for english. *Commun. ACM*, 38(11):39–41, 1995.

[OH01] V. Orengo and C. Huyck. A stemming algorithm for the Portuguese language. In *Proceedings of the 8th International Symposium on String Processing and Information Retrieval(SPIRE) 2001*, pages 186–193, 2001.

[Por80] M. F. Porter. An algorithm for suffix stripping. In *ACM SIGIR Conference on R&D in Information Retrieval*, pages 318–327, 1980. A plain text version of the article and implementations of the proposed algorithm can be found in: http://www.tartarus.org/~martin/PorterStemmer/.

[Pro04] Protégé–2000, 2004. http://protege.stanford.edu.

[SWM04] M. Smith, C. Welthy, and D. McGuiness. OWL Web Ontology Language Guide. Technical report, World Wide Web Consortium, 2004. This technical report can be found in: http://www.w3.org/TR/2004/REC-owl-guide-20040210/.

[Sys04] Sun Systems. Java Media Framework API, 2004. http://java.sun.com/products/java-media/jmf/.

[Ued02] R. Ueda. Ispell Dictionary for Brazilian Portuguese: br.ispell, 2002. http://www.ime.usp.br/ ueda/br.ispell/.

[vR79] C. J. van Rijsbergen. *Information Retrieval.* Butterworths, 2nd edition, 1979. An electronic version of this book can be found in: http://citeseer.ist.psu.edu/vanrijsbergen79information.html.

[WBvG+02] M. Worring, A. Bagdanov, J. v. Gemert, J-M. Geusebroek, M. Hoang, A.Th. Schreiber, C.G.M. Snoek, J. Vendrig, J. Wielemaker, and A.W.M. Smeulders. Interactive indexing and retrieval of multimedia content. In *Proceedings of the 29th Annual Conference on Current Trends in Theory and Practice of Informatics (SOFSEM), volume 2540 of Lecture Notes in Computer Science*, pages 135–148, Milovy, Czech Republic, 2002. Springer-Verlag.

This article was processed using the LaTeX macro package with LLNCS style