INSTITUT NATIONAL DE RECHERCHE EN INFORMATIQUE ET EN AUTOMATIQUE

Project-Team TEXMEX

Efficient Exploitation of Multimedia Documents: Exploring, Indexing and Searching in Very Large Databases

Rennes - Bretagne-Atlantique

THEME SYM

Activity Report

2008

Table of contents

TEXMEX *is a common project-team with CNRS, University of Rennes 1 and INSA. It has been created on January the 1st, 2002 and became an INRIA project on November the 1st, 2002.*

1. Team

Research Scientist

Patrick Gros [Team Leader, Senior Research Scientist, INRIA, HdR]
Laurent Amsaleg [Research Scientist, CNRS]
Vincent Claveau [Research Scientist, CNRS]

Faculty Member

Stéphane Huet [Assistant Professor, Univ. Rennes 1, until August 31th]
Ewa Kijak [Associate Professor, Univ Rennes 1]
Annie Morin [Associate Professor, Univ Rennes 1, HdR]
François Poulet [Associate Professor, Univ Rennes 1, HdR]
Pascale Sébillot [Professor, INSA, HdR]

External Collaborator

Emmanuelle Martienne [Associate Professor, Univ. Rennes 2]
Fabienne Moreau [Associate Professor, Univ. Rennes 2]
Laurent Ughetto [Associate Professor, Univ. Rennes 2]

Technical Staff

Mathieu Ben [INRIA Technical Staff, since October 1th]
Sébastien Campion [INRIA Research Engineer]
Arnaud Dupuis [INRIA Technical Staff, until November 30th]
Stacy Payne [INRIA Technical Staff, also with SAF, since September 1th]

PhD Student

Siwar Baghdadi [CIFRE grant with Thomson]
Camille Guinaudeau [Quaero project-Brittany Council grant, since October 1th]
Gwénolé Lecorvé [MESR grant]
Gaël Manson [CIFRE grant with Orange]
Pham Nguyen Khang [also with *Institut francophone pour l'informatique*, Hanoï, Vietnam]
Romain Tavenard [ENS Cachan grant]
Pierre Tirilly [CNRS - Brittany Council grant]
Joaquin Zepeda [ICOS-HD project, also with TEMICS]

Post-Doctoral Fellow

Ariane Herbulot [ICOS-HD project, also with TEMICS, until August 31th]
Zein Al Abidin Ibrahim [Semim@ges project, since March 1th]

Visiting Scientist

Kári Harðarson [Lecturer at Reykjavík University, from August 28th to January 3rd 2009]

Administrative Assistant

Loïc Lesage [Secretary INRIA, partial position in the project-team]

2. Overall Objectives

2.1. Overall Objectives

Keywords: *data analysis*, *databases*, *documents*, *image recognition*, *indexing*, *machine learning*, *multimedia*, *natural language processing*, *search*, *television*, *video*.

With the success of sites like Youtube or DailyMotion, with the development of the Digital Terrestrial TV, it is now obvious that the digital videos have invaded our usual information channels like the web. While such new documents are now available in huge quantities, using them remains difficult. Beyond the storage problem, they are not easy to manipulate, browse, describe, search, summarize, visualize as soon as the simple scenario "1. search the title by keywords 2. watch the complete document" does not fulfill the user's needs anymore. That is, in most cases.

Most usages are linked with the key concept of repurposing. Videos are a raw material that each user recombines in a new way, to offer new views of the content, to adapt it to new devices (ranging from HD TV sets to mobile phones), to mix it with other videos, to answer information queries... Somehow, each use of a video gives raise to a new short-lived document that exists only while it is viewed. Achieving such a repurposing process implies the ability to manipulate videos extracts as easily as words in a text.

Many applications exist in both professional and domestic areas. On the professional side, such applications include transforming a TV broadcast program into a web site, a DVD or a mobile phone service, switching from a traditional TV program to an interactive one, better exploiting TV and video archives, constructing new video services (video on demand, video edition...). On the domestic side, video summarizing can be of great help, as can a better management of the videos locally recorded, or simple tools to face the exponential number of TV channels available that increase the quantity of interesting documents available, overall increasing but make them really hard to find.

In order to face such new application needs, we propose a multi-field work, gathering in a single team specialists that are able to deal with the various media and aspects of large video collections: image, video, text, sound and speech, but also data analysis, indexing, machine learning... The main goal of this work is to segment, structure, describe, or delinearize the multimedia content in order to be able to recombine or re-use that content in new conditions. The focus on the document analysis aspect of the problem is an explicit choice since it is the first mandatory step of any subsequent application, but using the descriptions obtained by the processing tools we develop is also an important goal of our activity.

To summarize our research project in one short sentence, let us say that we would like our PCs to be able to watch TV and use what has been watched and understood in new innovative services. The main challenges to address in order to reach that goal are: the size of the documents and of the document collections to be processed, the necessity to process several media in a joint manner and to obtain a high level of semantics, the variety of contents, of contexts, of needs and usages, linked to the difficulty to manage such documents on a traditional interface.

Our own research is organized in three directions: 1- developing advanced algorithms of data analysis, description and indexing, 2- searching new techniques for linguistic information acquisition and use, 3- building new processing tools for audiovisual documents.

2.1.1. Advanced Algorithms of Data Analysis, Description and Indexing

Processing multimedia documents produces most of the time lots of descriptive metadata. These metadata can take many different aspects ranging from a simple label issued from a limited list, to high dimensional vectors or matrices of any kind; they can be numeric or symbolic, exact, approximate or noisy... As examples, image descriptors are usually vectors whose dimension can vary between 2 and 900, while text descriptors are vectors of much higher dimension, up to 100,000 but that are very sparse. Real size collections of documents can produce sets of billions of such vectors.

Most of the operations to be achieved on the documents are in fact translated in terms of operations on their metadata and the former appears as key objects to be manipulated. Although their nature is much simpler than the data used to compute them, these metadata require specific tools and algorithms to cope with their particular structure and volume. Our work concerns mainly three domains:

- data analysis techniques, eventually coupled to data visualization techniques, to study the structure of large sets of metadata, with applications to classical problems like data classification, clustering, sampling, or modeling,

- advanced data indexing techniques in order to speed-up the manipulation of these metadata for retrieval or query answering problems,
- description of compressed, watermarked or attacked data.

2.1.2. New Techniques for Linguistic Information Acquisition and Use

Natural languages are a privileged way to carry high level semantic information. Used in speech from an audio track, in textual format or overlaid in images or videos, alone or associated with images, graphics or tables, organized linearly or with hyperlink, expressed in English, French, or Chinese, this linguistic information may take many different forms, but always exhibits a common basic structure: it is composed of sequences of words. Building techniques that preserve the subtle links existing between these words, their representations with letters or other symbols and the semantics they carry is a difficult challenge.

As an example, actual search engines work at the representation level (they search sequences of letters), and do not consider the meaning of the searched words. Therefore, they do not use the fact that “bike” and “bicycle” represent a single concept while “bank” has at least two different meanings (a river bank and a financial institution).

Extracting high level information is the goal of our work. First, acquisition techniques that allow us to associate pieces of semantics with words, to create links between words are still an active field of research. Once this linguistic information is available, its use opens new problems. For example, in search engines, new pieces of information can be stored and the representation of the data can be improved in order to increase the quality of the results.

2.1.3. New Processing Tools for Audiovisual Documents

Audiovisual documents have the main characteristic to be temporal documents. One of the consequences is that they cannot be watched or listened to globally, but only by a linear process that takes some time. On the processing side, these documents often mix several media (image track, sound track, some text) that should be all taken into account to understand the meaning and the structure of the document. They can also have an endless stream structure with no clear temporal boundaries, like on most TV or radio channels. Therefore, there is an important need to segment and structure them, at various scales, before describing the pieces that are obtained.

Our work is organized in three directions. Segmenting and structuring long TV streams (up to several weeks, 24 hours a day) is a first goal that allows to extract program and non program segments in these streams. These programs can then be structured at a finer level. Finally, once the structure is extracted, we use the linguistic information to describe and characterize the various segments. In all this work, the interaction between the various media is a constant source of difficulty, but also of inspiration.

2.2. Highlights of the Year

In 2008, three events in TEXMEX are worth noting:

1. The paper "A Language Modeling Approach to Image Classification " by Pierre Tirilly, Vincent Claveau, and Patrick Gros obtained the best paper award of the NEM summit 2008 which took place from October 13^{th} to 15^{th} in Saint Malo, France.
2. Our associate team with Reykjavìk University gave raise to a newborn company, Eff2 (http://www.eff2.net), to commercialize the copy detection technology developed in the team.
3. The technology developed for our video indexing platform will be commercialized by another company, Powedia (http://www.powedia.com), which was awarded by a prize from the French Ministry of Research (see Section 7.3.1).
4. The Quaero project eventually started in May 2008. This 5-year research program about multimedia description involves many members of the team working on key topics for TEXMEX (see Section 7.1.4).

3. Scientific Foundations

3.1. Background

This section presents some of the basic techniques used in the team. They are basic blocks used in many research projects as well as research topics by themselves, since we usually use them in new contexts.

3.2. Local Image Description

Keywords: *Image Description*, *Local Descriptors*, *Metadata*.

In most contexts where images are to be compared, a direct comparison is impossible. Images are compressed in different formats, most formats are error-prone, images can be re-sized, cropped... The solution consists in computing descriptors from the images and to turn image comparison into descriptor comparison. This can be done if, on the one hand, the descriptors contain some information on the image content, while, on the other hand, they do not depend on the image format, size or on transformations the image can undergo.

The most classical method associates a unique global descriptor with each image, e.g. a color histogram or correlogram, a texture descriptor. Such descriptors are easy to compute and use, but they usually fail to handle cropping and cannot be used for object recognition. A second method consists in extracting regions in the image and to associate a descriptor with each of these regions. Most of the time, this is done by extracting points (called interest points) with a Harris-like detector [72], and by considering a circular or elliptic region around each of these points.

The differential invariants were among the first local descriptors used. Established by Florack [69], their use for image comparison was proposed by Schmid [86]. Each descriptor is a combination of the first derivatives of the signal at the interest point. These descriptors appeared experimentally to be very robust to geometric and photometric transforms. An even more powerful descriptor was then proposed by Lowe: the SIFT descriptor [75]. It is composed of 16 local histograms of gradient directions around the interest point.

Local descriptors can be used in many applications: image comparison for object recognition, image copy detection, detection of repeats in television streams... While they are very reliable, local descriptors are not without problems. As many descriptors can be computed for a single image, a collection of one million images can generate a database of one billion of descriptors. That is why specific indexing techniques are required. Up to now, most of them are computed from decompressed images, while most formats images are stored compressed. Thus it would be interesting to directly compute the descriptors in the compressed domain. Finally, their evaluation for very large image collection (several millions of images) is still an open and interesting problem.

3.3. Corpus-based Text Description and Machine Learning

Keywords: *Machine-learning*, *Text Description*.

Our work on textual material (textual documents, transcriptions of speech documents, captions in images or videos, etc.) is characterized by a chiefly corpus-based approach, as opposed to an introspective one. A corpus is for us a huge collection of textual documents, gathered or used for a precise objective. We thus exploit specialized (abstract of biomedical articles, computer science texts, etc.) or non specialized (newspapers, broadcast news, etc.) collections for our various studies. In TEXMEX, according to our applications, different kinds of knowledge can be extracted from the textual material. For example, we automatically extract terms characteristic of each successive topic in a corpus with no a priori knowledge; we produce representations for documents in an indexing perspective [85]; we acquire lexical resources from the collections (morphological families, semantic relations, translation equivalences, etc.) in order to better grasp relations between segments of texts in which a same idea is expressed with different terms or in different languages...

In the domain of the corpus-based text processing, many researches have been undergone in the last decade. While most of them are essentially based on statistical methods, symbolic approaches also present a growing interest [60]. For our various problems involving language processing, we use both approaches, making the most of existing machine learning techniques or proposing new ones. Relying on advantages of both methods, we aim at developing machine learning solutions that are automatic and generic enough to make it possible to extract, from a corpus, the kind of elements required by a given task.

3.4. Stochastic Models for Multimodal Analysis

Keywords: *Image Stochastic Models*, *Multimedia*.

Describing multimedia documents, *i.e.* documents that contain several modalities (e.g. text, images, sound) requires to take all these modalities into account since these modalities can contain complementary pieces of information. The problem is that the various modalities are only weakly synchronized, they do not have the same rate and combining the information that can be extracted from them is not obvious. Of course, we would like to find generic ways to combine these pieces of information. Stochastic models appear as a well dedicated tool for such combinations, especially for image and sound information.

Markov models are composed of a set of states, of transition probabilities between these states and of emission probabilities that provide the probability to emit a given symbol at a given state. Such models allow to generate sequences. Starting from an initial state, they iteratively emit a symbol and then switch in a subsequent state according to the respective probability distributions. These models can be used in an indirect way. Given a sequence of symbols (called observations), hidden Markov models (HMM, [81]) aims at finding the best sequence of states that can explain this sequence. The Viterbi algorithm provides an optimal solution to this problem.

For such HMM, the structure and probability distributions need to be a priori determined. They can be fixed manually (this is the case for the structure: number of states and their topology), or estimated from example data (this is often the case for the probability distributions). Given a document, such an HMM can be used to retrieve its structure from the features that can be extracted. As a matter of fact, these models allow an audiovisual analysis of the videos, the symbols being composed of a video and an audio component.

Two of the main drawbacks of the HMM is that they can only emit a unique symbol per state, and that they imply that the duration in a given state follows an exponential distribution. Such drawbacks can be circumvented by segment models [79]. These models are an extension of HMM were each state can emit several symbols and contains a duration model that governs the number of symbols emitted (or observed) for this state. Such a scheme allows us to process features at different rates.

Bayesian networks are an even more general model family. Static Bayesian networks [67] are composed of a set of random variables linked by edges indicating their conditional dependency. Such models allow us to learn from example data the distributions and links between the variables. A key point is that both the network structure and the distributions of the variables can be learned. As such, these networks are difficult to use in the case of temporal phenomenon.

Dynamic Bayesian [76] networks are a generalization of the previous models. Such networks are composed of an elementary network that is replicated at each time stamp. Duration variable can be added in order to provide some flexibility on the time processing, like it was the case with segment models.

While HMM and segment models are well suited for dense segmentation of video streams, Bayesian networks offer better capabilities for sparse event detection. Defining a trash state that correspond to non event segment is a well known problem is speech recognition: computing the observation probabilities in such a state is very difficult.

3.5. Multidimensional Indexing Techniques

Keywords: *Approximate Searches*, *Curse of Dimensionality*, *Databases*, *Multidimensional Indexing Techniques*, *Nearest Neighbors*.

Techniques for indexing multimedia data are needed to preserve the efficiency of search processes as soon as the data to search in becomes large in volume and/or in dimension. These techniques aim at reducing the number of I/Os and CPU cycles needed to perform a search. Two classes of multi-dimensional indexing methods can be distinguished: exact nearest neighbor (NN) searches and approximate NN-search schemes.

Traditional multidimensional indexing techniques typically divide the data space into cells containing vectors [64]. Cell construction strategies can be classified in two broad categories: *data-partitioning* indexing methods that divide the data space according to the distribution of data, and *space-partitioning* indexing methods that divide the data space along predefined lines and store each descriptor in the appropriate cell. NN-algorithms typically use the geometrical properties of (minimum bounding) cells to eliminate cells that cannot have any impact on the result of the current query [66].

Many data-partitioning index methods derive from the seminal R-Tree [71], and their differences lie in the properties of the shapes used to build cells and/or in the degree of overlapping between cells. Well known space-partitioning techniques are somehow related to the K-D-B-Tree [83], and differ on the way space is split and cells encoded.

Unfortunately, the "curse of dimensionality" phenomenon makes these traditional approaches ineffective in high-dimensional spaces [62]. This phenomenon is particularly prevalent when performing *exact* NN-searches. There is therefore an increasing interest in performing *approximate* NN-searches, where result quality is traded for reduced query execution time. Many approaches to approximate NN-searches have been published; their description can be found in [62].

Some approaches simply rely on dimensionality reduction techniques, such as PCA, but their use remains problematic when facing very high-dimensional datasets. Other approaches abort the search process early, after having accessed an arbitrary and predetermined number of cells. While this is highly effective, it does not give any clue on the quality of the result returned to the user. Some other approaches consider an approximation of the sizes of cells instead of considering their exact sizes, making somehow cells "smaller". Shrunk cells increase efficiency of retrievals as they reduce overlap in space, but interesting vectors might be missed, however.

Recently, several approaches have transformed costly nearest neighbor searches in multidimensional space into efficient uni-dimensional accesses. One approach using locality sensitive hashing (LSH) techniques [70] uses several hash functions such that co-located vectors are likely to collide in buckets. Fagin *et al.* [68] proposed a framework based on projecting the descriptors onto a limited set of random lines, each line giving a ranking of the database descriptors with respect to the query descriptor.

3.6. Data Mining Methods

Data Mining (DM) is the core of KDD (Knowledge Discovery in Databases) whatever the contents of the databases are. Here, we focus on some aspects of DM we use to describe documents and to retrieve information. There are two major goals to DM: description and prediction. The descriptive part includes unsupervised and visualization aspects while prediction is often referred to as supervised mining.

The description step very often includes feature extraction and dimensional reduction. As we deal mainly with contingency tables crossing "documents and words", we intensively use factorial correspondence analysis. "Documents" in this context can be a text as well as an image.

Correspondence analysis is a descriptive/exploratory technique designed to analyze simple two-way and multi-way tables containing some measure of correspondence between the rows and columns. The results provide information which is similar in nature to those produced by factor analysis techniques, and they allow one to explore the structure of categorical variables included in the table. The most common kind of table of this type is the two-way frequency cross-tabulation table. There are several parallels in interpretation between correspondence analysis and factor analysis: suppose one could find a lower-dimensional space, in which to position the row points in a manner that retains all, or almost all, of the information about the differences between the rows. One could then present all information about the similarities between the rows in a simple

1, 2, or 3-dimensional graph. The presentation and interpretation of very large tables could greatly benefit from the simplification that can be achieved via correspondence analysis (CA).

One of the most important concept in CA is inertia, *i.e.* the dispersion of either row points or column points around their gravity center. The inertia is linked to the total Pearson χ^2 for the two-way table. Some rows and/or some columns will be more important due to their quality in a reduced dimensional space and their relative inertia. The quality of a point represents the proportion of the contribution of that point to the overall inertia that can be accounted for by the chosen number of dimensions. However, it does not indicate whether or not, and to what extent, the respective point does in fact contribute to the overall inertia (χ^2 value). The relative inertia represents the proportion of the total inertia accounted for by the respective point, and it is independent of the number of dimensions chosen by the user. We use the relative inertia and quality of points to characterize clusters of documents. The outputs of CA are generally very large. At this step, we use different visualization methods to focus on the most important results of the analysis.

In the supervised classification task, a lot of algorithms can be used, the most popular ones are the decision trees and more recently the Support Vector Machines (SVM). SVMs provide very good results in supervised classification but they are used as "black boxes" (their results are difficult to explain). We use graphical methods to help the user understanding the SVM results, based on the data distribution according to the distance to the separating boundary computed by the SVM and another visualization method (like scatter matrices or parallel coordinates) to try to explain this boundary. Other drawbacks of SVM algorithms are their computational cost and large memory requirement to deal with very large datasets. We have developed a set of incremental and parallel SVM algorithms to classify very large datasets on standard computers.

4. Application Domains

4.1. Copyright Protection of Images

Keywords: *Copyright*, *Digital Pictures*, *IPR*, *Image Databases*, *Movie Makers*, *Photo Agencies*.

With the proliferation of high-speed Internet access, piracy of multimedia data has developed into a major problem and media distributors, such as photo agencies, are making strong efforts to protect their digital property. Today, many photo agencies expose their collections on the web with a view to selling access to the images. They typically create web pages of thumbnails, from which it is possible to purchase high-resolution images that can be used for professional publications. Enforcing intellectual property rights and fighting against copyright violations is particularly important for these agencies as these images are a key source of revenue. The most problematic cases, and the ones that induce the largest losses, occur when “pirates” steal the images that are available on the Web and then make money by illegally reselling those images.

This applies to photo agencies, and also to producers of videos and movies. Despite the poor image quality, thousands of (low-resolution) videos are uploaded every day to video-sharing sites such as YouTube, eDonkey or BitTorrent. In 2005, a study conducted by the Motion Picture Association of America was published, which estimated that their members lost 2,3 billion US$ in sales due to video piracy over the Internet. Due to the high risk of piracy, movie producers have tried many means to restrict illegal distribution of their material, albeit with very limited success.

Photo and video pirates have found many ways to circumvent even the most clever protection mechanisms. In order to cover up their tracks, stolen photos are typically cropped, scaled, their colors are slightly modified; videos, once ripped, are typically compressed, modified and re-encoded, making them more suitable for easy downloading. Another very popular method for stealing videos is camcording, where pirates smuggle digital camcorders into a movie theater and record what is projected on the screen. Once back home, that goes to the web.

Clearly, this environment calls for an automatic content-based copyright enforcement system, for images, videos, and also audio as music gets heavily pirated. Such a system needs to be effective as it must cope with often severe attacks against the contents to protect, and efficient as it must rapidly spot the original contents from a huge reference collection.

4.2. Video Database Management

Keywords: *Multimedia*, *Video Bases*, *Video Structuring*.

The existing video databases are generally little digitized. The progressive migration to digital television should quickly change this point. As a matter of fact, the French TV channel TF1 switched to an entirely digitized production, the cameras remaining the only analogical spot. Treatment, assembly and diffusion are digital. In addition, domestic digital decoders can, from now on, be equipped with hard disks allowing a storage initially modest, of ten hours of video, but larger in the long term, of a thousand of hours.

One can distinguish two types of digital files: private and professional files. On one hand, the files of private individuals include recordings of broadcasted programs and films recorded using digital camcorders. It is unlikely that users will rigorously manage such collections; thus, there is a great need for tools to help the user: automatic creation of summaries and synopses to allow finding information easily or to have within few minutes a general idea of a program. Even if the service is rustic, it is initially evaluated according to the added-value brought to a system (video tape recorder, decoder), must remain not very expensive, but will benefit from a large diffusion.

On the other hand, these are professional files: TV channel archives, cineclubs, producers... These files are of a much larger size, but benefit from the attentive care of professionals of documentation and archiving. In this field, the systems can be much more expensive and are judged according to the profits of productivity and the assistance which they bring to archivists, journalists and users.

A crucial problem for many professionals is the need to produce documents in many formats for various terminals from the same raw material without multiplying the editing costs. The aim of such a *repurposing* is for example to produce a DVD, a web site or an alert service by mobile phone from a TV program at the minimum cost. The basic idea is to describe the documents in such a way that they can be easily manipulated and reconfigured easily.

4.3. Textual Database Management

Keywords: *Bibliography*, *Indexing*, *Natural Language Processing*.

Searching in large textual corpora has already been the topic of many researches. The current stakes are the management of very large volumes of data, the possibility to answer requests relating more on concepts than on simple inclusions of words in the texts, and the characterization of sets of texts.

We work on the exploitation of scientific bibliographical bases. The explosion of the number of scientific publications makes the retrieval of relevant data for a researcher a very difficult task. The generalization of document indexing in data banks did not solve the problem. The main difficulty is to choose the keywords which will encircle a domain of interest. The statistical method used, the factorial analysis of correspondences, makes it possible to index the documents or a whole set of documents and to provide the list of the most discriminating keywords for these documents. The index validation is carried out by searching information in a database more general than the one used to build the index and by studying the retrieved documents. That in general makes it possible to still reduce the subset of words characterizing a field.

We also explore scientific documentary corpora to solve two different problems: to index the publications by the way of meta-keys and to identify the relevant publications in a large textual database. For that, we use factorial data analysis which allows us to find the minimal sets of relevant words that we call meta-keys and to free the bibliographical search from the problems of noise and silence. The performances of factorial correspondence analysis are sharply greater than classic search by logical equation.

5. Software

5.1. Software

Several software programs have been developed in the team over the years: I-DESCRIPTION (APP deposit number: IDDN.FR.001.270047.000.S.P.2003.000.21000), ASARES (IDDN.FR.001.0032.000.S.C.2005.000.20900), ANAMORPHO (IDDN.FR.001.050022.000.S.P.2008.000.20900), FAESTOS (IDDN.FR.001.470029.000.S.P.2006.000.40000), 2PAC (IDDN.FR.001.470028.000.S.P.2006.000.40000), NAVITEX (IDDN.FR.001.190034.000.S.P.2007.000.40000). Please refer to the previous activity reports for a description of these programs. In 2008, the following pieces of software were developed or updated.

5.1.1. *DiVATex*

Participants: Sébastien Campion [correspondant], Arnaud Dupuis, Patrick Gros.

The library DIVATEX (*Di*stant *V*ideo *A*ccess *Tex*mex) is the evolution of a previous library called TMX-DIVA-SOLUTION. This software enables an easy access to a collection of audio, video and picture contents stored on a remote machine. It contains a server on the remote machine, in charge of extracting from content files (video, audio or picture) the image or sound data needed by the user, and of sending it to the client. The client part of the software decodes what it receives from the server and provides the user with the exact data needed (the exact image or sound information). The first version of this software (TMX-DIVA-SOLUTION) was deposited with the "Agence pour la Protection des Programmes" (APP) under the number IDDN.FR.001.320006.000.S.P.2006.000.40000. The new version (DIVATEX), in which several bug were fixed and accessibility on local content was improved, is currently in a process of being deposited with the APP.

5.1.2. *Telemex*

Participants: Sébastien Campion [correspondant], Arnaud Dupuis, Patrick Gros.

The TELEMEX web interface was made to manage huge audiovisual corpora whatever the user's operating system. Using this interface, each authorized user can record TV programs, visualize it and navigate inside video files. Each video can be associated with a TV-anytime file, generated by the software NAVITEX. For each video sequence defined by TV-anytime files, TELEMEX generates a unique identifier. Then, this video can be used by DIVATEX. The new version (TELEMEX) is currently in a process of being deposited with the APP. This new version was adapted to the digital television and webradios.

5.2. Experimental Platform

Participants: Laurent Amsaleg, Mathieu Ben, Sébastien Campion [correspondant], Arnaud Dupuis, Patrick Gros, Pascale Sébillot.

Until 2005, we used various computers to store our data and to carry out our experiments. In 2005, we began some work to specify and set-up dedicated equipment to experiment on very large collections of data. During 2006 and 2007, we specified, bought and installed our first complete platform. It is organized around a very large storage capacity (up to 70TB), and contains 4 acquisition devices (for Digital Terrestrial TV), 3 video servers, and 15 computing servers partially included in the local cluster architecture.

In addition to this, three pieces of software were developed for this experimental framework: one to manage the acquisition process and all the data already stored (TELEMEX), the second one to annotate audiovisual corpora (NAVITEX) stored on the server, and the third one to ensure an easy access to the data from processing programs DIVATEX (video, audio and pictures). The platform will be completed with dedicated software to manage all the metadata associated with the data. This platform is funded by a joint effort of INRIA, INSA Rennes and University of Rennes 1.

In 2008, we began to build up a corpus of multimedia data. It consists in a continuous recording (6 months) of two TV channels and three radios. It also includes web pages related to these contents captured on broadcaster's website. This corpus is expected to be used for different studies like the treatment of news along the time and to provide sub-corpus like TV news. Within the Quaero project (see below), we also plan to annotate manually all the TV content.

6. New Results

6.1. Advanced Algorithms of Data Analysis, Description and Indexing

6.1.1. Advanced Image Description Techniques

6.1.1.1. Image Joint Description and Compression

Keywords: *Image Compression*, *Image Description*, *Image Indexing*.

Participants: Ariane Herbulot, Ewa Kijak, Joaquin Zepeda.

This is a joint work with the TEMICS *project-team (C. Guillemot).*

The objective of the study initiated in 2007 is to design scalable signal representation and approximation methods amenable to both compression (that is with sparseness properties) and description. Image representations obtained with various transforms, e.g., separable wavelet transforms, curvelets and bandlets have been considered for compression and de-noising applications. However, these critically-sampled transforms do not allow the extraction of low level signal features (points, edges, ridges, blobs) or of local descriptors. Feature extraction requires the image representation to be covariant under a set of admissible transformations. On the other hand, the Gaussian scale space is often used for description, however it is not amenable to compression. We thus investigate sparse representations methods for local image description. Sparse representations allow the use of inverted files, which provide a solution to the indexation of high dimensional data (specifically, textual documents) by taking advantage of sparse vectors properties. Document similarity calculations are thus carried out efficiently using the scalar products between these sparse vectors. In this context, two approaches are currently being considered: descriptors sparse representation and signal sparse descriptor.

One robust (non-sparse) descriptor presenting indexing issues due to its large dimensionality is the SIFT descriptor. A recent approach referred to as *Video Google* tackles the problem by forming a single sparse descriptor obtained from multiple input non-sparse SIFT descriptors. The approach consists in vector quantizing each SIFT descriptor and then taking a (weighted) histogram of codeword indices. We proposed a related approach that applies a pursuit-based sparse decomposition to each SIFT descriptor to obtain a sparse vector for each input SIFT descriptor, thus retaining the local characteristic of the input descriptors rather than forming a single global descriptor. The descriptors still enable the use of inverted file type indices.

Video Google's synthesis of a sparse SIFT vector begs the question of whether one can come up with a descriptor that is sparse by design, thus avoiding the performance loss due to the sparse synthesis while retaining the search benefits related to sparsity. Our aim is to adapt existing work in sparse decomposition (designed with compression and prediction in mind) to the construction of image descriptors displaying covariance to the set of admissible transformations. In this context we have identified three problems to address: 1) dictionary design, 2) atom selection method and 3) descriptor comparison method. Regarding dictionary design, the set of atoms comprising the dictionary should be closed under the set of admissible transformations. For atom selection we propose a new non-iterative method that select atoms that locally maximize the correlation between the dictionary function and the image patch in question. Such a method yields atoms that vary predictably with the transformations from the set of admissible image transformations. These predictable variations are at the root of the third problem to address, as descriptor comparison methods must take these variations into account when calculating descriptor similarities. One current approach considered of accounting for these variations in atom selection is to compare atoms using, for example, a distance defined in the transformation space, and to include this atom similarity measure in the descriptor distance expression.

Another aspect of this work is to study descriptions able to deal with the various resolution of images provided by scalable coding. Through the post-doctoral position of Ariane Herbulot, we studied an adaptation of SIFT descriptors to the context of scalable coding. It results in a new representation of SIFT descriptors, more suitable for video copy detection in the context of scalable coding.

6.1.1.2. NLP techniques for Image Description

Keywords: *Image-Text Interaction*, *NLP approaches for Image Description*.

Participants: Vincent Claveau, Patrick Gros, Pierre Tirilly.

Our work is based on the *video-google* [88] retrieval scheme for images. This approach describes pictures as sets of local image features called visual words. This description is more semantic than the classical global features formerly used in image retrieval and classification such as color histograms or Gabor filters. Moreover, it provides a bag-of-visual words description of images, very close to the traditional bag-of-words representation of textual documents used in information retrieval [90]. Describing images as text gives us the opportunity to adapt many text retrieval and classification techniques to the field of image retrieval and classification.

Among these techniques, we worked on the following in 2008:

- probabilistic Latent Semantic Indexing (pLSA) [73]: we showed that pLSA can be used to eliminate about one third of the visual words describing a collection, allowing to perform faster image matching. In the case of image classification, this method yields better results than the use of the complete vocabulary.
- language models [63]: language models can perform efficient and effective image classification by taking into account sequences of words instead of independent words. Since the geometrical layout of visual words in an image is related to the meaning of this image, we used language models to perform image classification so that this layout can be exploited. Images are first described as sequences (called visual sentences) by simply projecting visual words on a well-chosen axis. Then language models are used to classify the visual sentences. Experiments showed that taking the spatial information of the visual words into account to perform image classification yields better results than relying on independent visual words. They also showed that this approach yields better classification results than a state-of-the-art SVM classifier. Our conference paper on this subject was awarded a best-paper prize at the NEM'08 conference.
- weighting schemes for image retrieval: former bag-of-visual words approaches used the standard *tf.idf* weighting scheme to enhance image retrieval performances [88]. However, many other weighting schemes exist and had not been used in the case of image retrieval yet. We therefore tested several vector-space-based [84], [65] and probabilistic-based [82], [61] weighting schemes in the case of image retrieval. Experiments showed that some weighting schemes can slightly improve image retrieval, although none of these yields the best results for all datasets. We also showed that the L1 distance performs better than the other usual distances, L2 and cosine.

6.1.2. Advanced Data Analysis Techniques

6.1.2.1. Intensive Use of Factorial Analysis for Text and Textual Streams Mining

Keywords: *Correspondence Analysis*, *Visualization*.

Participant: Annie Morin.

Textual data can be easily transformed in frequency tables and any method working on contingency tables can be used to process them. Besides, with the important amount of available textual data, we need to find convenient ways to process the data and to get invaluable information. It appears that the use of factorial correspondence analysis allows us to get most of the information included in the data. But even after the data processing, we still have a big amount of material and we need visualization tools to display it. We study the relevance of different indicators used to cluster the words on one side and the documents on the other side and we are concerned by the visualization of the outputs of factorial analysis: we need to help the user to go through the huge amount of information we get and to select the most relevant points. Most of the time, we do not pre-process the texts: that means that there is no lemmatization. We also start exploring temporal changes in textual data and the first experiments have been done on newspaper corpus from 1987 to 2003. For the moment, we mainly focus on the visualization of results.

6.1.2.2. Intensive Use of SVM for Text Mining and Image Mining

Keywords: *Parallel and Distributed Algorithms*, *Support Vector Machine*.

Participants: Pham Nguyen Khang, François Poulet.

Support Vector Machines (SVM) and kernel methods are known to provide accurate models but the learning task usually needs a quadratic program, so this task for very large datasets requires a large memory capacity and a long time. We have developed new algorithms: a boosting of least squares SVM to classify very large datasets on standard personal computers and incremental and parallel SVMs. The incremental part of the algorithm avoids us to load the whole dataset in main memory; we only need to have a small part of the dataset in main memory to build a part of the data model. Then we put together the partial models to get the full one with the same accuracy as usual algorithm; it solves the memory capacity problem of SVM algorithms.

To solve the computational time problem we have distributed the computation of the data blocks on different computers by the way of parallel and distributed algorithms. The first versions of the algorithms were based on a CPU distributed software program, then we have used GP-GPU (General Purpose GPU) versions to significantly improve the algorithm speed [18]. The time needed for usual SVM algorithms like libSVM is divided by at least 1000.

We have extended the least squares SVM algorithm (LS-SVM). The first step was to adapt the algorithm to deal with datasets having a very large number of dimensions (like in text or image mining). Then we have applied boosting to LS-SVM for mining huge datasets having simultaneously a very large number of datapoints and dimensions on standard computers [27]. The performance of the new algorithm has been evaluated on large datasets from Machine Learning repository like Reuters-21578 or Forest Cover Type and image datasets. The accuracy is increased in almost all datasets compared to LibSVM.

6.1.2.3. Intensive Use of Factorial Analysis for Image Mining

Keywords: *Correspondence Analysis*, *Visualization*.

Participants: Patrick Gros, Annie Morin, Pham Nguyen Khang, François Poulet.

This work is done with Institut francophone pour l'informatique, Hanoï, Vietnam.

To analyze and retrieve information in image databases, we use the same method as in textual data analysis. This work is apart of the Ph.D. thesis of Pham Nguyen Khang. That means that in order to apply Correspondence Analysis (CA) to images, we must define "words" in images. This is usually achieved by 2 stages: (1) vector quantizing automatically extracted local image descriptors (*i.e.* SIFT) and (2) applying a clustering algorithm (*i.e.* k-mean clustering) on the set of descriptors to form "visual words". Once the visual words are defined, we then construct a contingency table by crossing visual words and images.

We began experiments by applying CA on the "Alogic" database (961 images). First, a vocabulary of 1000 visual words was computed from 60 first images using SIFT descriptors (code of D. Lowe) and a k-mean algorithm. CA was then applied on a contingency table of 961 x 1000. We kept only 30 first axes and used those axes for computing image similarity (Euclidean distance was used). Surprisingly, we found some groups of images which belong to the same categories (toys, houses, Eiffel tower).

Motivated by this promising result we continued our approach on "caltech4" database (4090 images of 5 categories: faces, airplane, motorbikes, cars (rear) and backgrounds) [87]. About 3000 descriptors sampled from all of descriptors (a third for every category) were clustered to form 2224 visual words. We explored this database on 2 tasks: image categorization and image retrieval. For the first task, we applied CA on the contingency table representing the database and kept only 7 axes (for comparison to PLSA trained with 7 hidden topics). A k-mean algorithm was then invoked to form clusters (categories). The result showed that CA performed slightly better than PLSA.

For image retrieval task, we compared our approach to PLSA and TF*IDF using L1 distance, L2 distance and cosine similarity. In the case of PLSA and CA, the retrieval was performed very fast because the problem dimension was reduced from 2224 to 7. For PLSA and CA, cosine similarity gave better result than L1 and L2 distance. The performances of CA and PLSA were equivalent and much better than that of TF*IDF.

We have also proposed a method for scaling up the problem using inverted files based on image representation quality on axes. Every inverted file was associated with the well represented images (on the axis to which the file belong). Given an image query, the search began by choosing the appropriate inverted files and intersecting those files. The similarity computation is done only on a subset of images resulting from the intersection of inverted files (about 1/5 to 1/8 of entire database). The performance was degraded only about 0.3% with respect to the exhaustive method.

Besides, in order to reduce the learning time and to deal with large databases, we have developed an incremental version of CA algorithm which splits data in blocks and processes block after block. The parallelization of this algorithm on Graphic Process Unit (GPU) showed that the GPU version performs 20 to 30 times faster the CPU one. The retrieval's accuracy is also improved when combining contextual information into our index structure. For that, we have integrated the Contextual Dissimilarity Measure [74] into our retrieval platform using inverted files. This integration was explored in two directions: offline (correction terms are computed before the retrieval task) and online (correction terms are computed only on images in the candidate list). Tests realized on the Nister dataset [78] have shown a significant improvement of the accuracy. Our algorithm have been assessed for large scale datasets: we have merged the Nister dataset with 1 million other images. With our method, only 0.06% of dataset was explored (in 1/8 second) and that 0.06% contains 86.7% of relevant images.

6.1.3. Advanced Indexing Algorithms

6.1.3.1. Indexing for Very Large High Dimensional Spaces

Participant: Laurent Amsaleg.

This is a joint work with researchers from Reykjavík University.

We have highlighted in Section 4.1 the need for having automatic content-based image/video retrieval system for enforcing the protection of copyright. We have developed such a system that was evaluated against various data collections. In [12], we deeply investigated the behavior of our approach against a collection of 150,000 images. This paper gives a nice overview of our indexing schemes as well as the various options for creating the index.

We also did some larger scale experiments that are so far not yet documented. We ran the system over more than 1,000 hours of video and we did image recognition tests with about 10,000,000 images. So far, the our initial results tend to confirm that our indexing scheme scales up nicely, *i.e.*, its response time remains flat while the search quality does not severely degrade.

In general, while the query performance of high-dimensional indexing techniques has been well studied, however, the performance of index maintenance has not yet been reported in any detail. We therefore worked on studying the dynamic behavior of the NV-tree. The NV-tree has several configuration and implementation options that affect the performance of index maintenance. In [39] we report on an initial study of the effects of these options on the dynamic behavior of the NV-tree, and show that with appropriate implementation, significant performance improvements are observed.

6.1.3.2. Indexing with High-Dimensional Lattices

Participant: Laurent Amsaleg.

It is well known that doing nearest neighbors retrievals in high-dimensional spaces is a costly process. In addition to the research directions mentioned above, we have started investigating the use of Euclidean high-dimensional lattices for database indexing. Lattices have nice properties: they are spatial quantifiers, thereby generating a partition of the space and decoding (quantization step) can be done very quickly.

We have designed and implemented an initial indexing scheme that relies on lattices. We have investigated the pros and cons of using various types of lattices, namely the lattices Z^n, D_n^* and A_n^*, using data sets in their original feature space as well as indexing data after various dimensionality reductions.

It turns out that using lattices with no special care is problematic as it suffers from the dimensionality curse phenomenons. Nevertheless, we devised a hierarchical approach coping with sparse regions of space as well as dense regions. This proves to work well and we observed pretty nice results in terms of precision and recall [80].

During the curse of that research, we had to work on smart schemes to get as fast as possible the lattice cells that are near to the one where the query points falls. We discovered that this need was quite general and could potentially be beneficial to other applications not focusing on nearest neighbor retrievals.

Indeed, many applications might benefit from taking into account additional, well located, lattice points. Finding neighboring lattice points might allow to cope with unexpected severe noise in compression, by proposing at reconstruction time not only the best theoretical lattice point, but also other good candidates. Similarly, strong distortions in watermarking might be absorbed by considering several lattice points. Last, when the query image feature vector lies near a cell frontier, nearest neighbors might very well be located on the other side. Therefore fetching the contents of cells associated with neighboring lattice points typically improves the quality of the result.

Not all cells touching the one in which the query point lies bring useful information. The best cells are the ones that are the closest to the query point. Enumerating and then checking the distance to all possible lattice points around the query cell is impossible in practice as it gets exponential with the dimension.

We have proposed a very efficient approach to get the k lattice points that are the closest to a particular query point. That approach first determines the k faces of the query cell that are the closest to the query point. From these faces, it is trivial to get the lattice points lying behind. Our approach works for the lattices $\mathbb{Z}^n$, D_n^* and A_n^*. Its complexity is small: getting the k nearest lattice points requires far less operations than doing k decoding.

6.1.3.3. Vector Quantification and Hamming Spaces

Participants: Laurent Amsaleg, Patrick Gros, Ewa Kijak.

Content based image retrieval systems made a lot of progress lately. Their recognition power strongly improved due to local description schemes allowing for very good image matching despite severe distortions. In addition, the response time of image search engines dropped due to the invention of new indexing strategies such as the ones using visual words.

Unfortunately, querying a very large image collection still poses response time and accuracy problems. Briefly sketched, the best recognition techniques suffer from the crude approximations made by high-dimensional vector quantification techniques, and the fastest systems are slowed down due to the many distance calculations they perform.

We have started investigating this issue by building on top of one of the best state-of-the-art techniques in order to revise its fundamental principles. We devised initial indexing schemes to enhance the vector quantification phase by allowing for redundancy. In addition, we started exploring the impact of building a binary signature of quantified vectors allowing for the use of an ultra-fast Hamming-based filtering.

Very preliminary experiments were done using a collection of few millions descriptors. We are in the process of putting the conclusions together to decide on the next steps.

6.1.3.4. Security of Local Image Description Schemes

Participants: Laurent Amsaleg, Ewa Kijak.

This work is done in collaboration with Teddy Furon from the TEMICS *project-team.*

Techniques allowing for the copyright protection of images and/or videos are now mature. They all use some kind of low level description that is computed over the collection of images to protect. Then, the resulting set of descriptors is indexed in a database. When probed with the descriptions of a suspicious image, that database returns the most similar descriptors from which a confidence score is computed. That family of approaches works pretty well and copes with severe image modifications, i.e., original images are found even when the copy used as query is strongly degraded.

All existing schemes rely on local description schemes. They all find in images interest points. Each interest points is then used to compute an high-dimensional descriptor by analyzing the signal around the point. Descriptors are computed in such a way that they get invariant to many image distortions. That local scheme introduces in the image recognition process a great deal of redundancy, key to its increased robustness.

Like all the techniques related to security, it is normal to question its behavior in a non-cooperative settings. In this case, a pirate well aware of the description techniques and of the database indexing techniques can produce specific image distortions that could kill the effectiveness of the copyright enforcement system.

Typically, a pirate can try to play with the image contents such that interest points disappear, shift within the image, or appear in places where none were, such that image matching gets more difficult. Also, a pirate can twick the contents of specific images such that they always appear in the result set, thereby hiding the illegal copies because these specific images always get ranked first. Many other informed attacks can be devised.

We started an initial investigation of this problem by studying the way SIFT descriptors could be attacked by a pirate aware that this specific description technique is used. SIFT is made of two steps: the first step finds interest points and the second computes descriptors around found points. We therefore started our study with attacking the point extraction phase. We used a method based on local patch insertion which aims at either suppressing a point otherwise found, moving by one or more pixel a specific point or artificially forcing the extraction of a point otherwise ignored in the image. In each case, depending on the aggressiveness of the distortion to change points, we analyzed the visual artifact induced.

6.1.3.5. Describing Sequences for Audio/Video Retrieval

Participants: Laurent Amsaleg, Romain Tavenard.

Our work on this topic is done in close collaboration with Guillaume Gravier from the METISS *project-team.*

Today, we can quite well exploit rather large databases of still images and we know how to efficiently query them by content. The next step asks to turn our focus on more complex documents, typically video and audio. There are today several description techniques for audio and video but only very few techniques efficiently perform query-by-content on video or audio databases at large scale. Being able to use such techniques is particularly crucial for professional multimedia archivists.

The state of the art makes such searches possible, but only at a very small scale, i.e., on a very small amount of data. Today, no search technique is efficient enough to allow any practical usage of real-scale audio or video archives. In addition, simply extending existing multidimensional indexing techniques is not possible since they were designed for description schemes in which the concept of sequence is lacking.

We have started investigating this issue in 2007. Overall, deciding whether two sequences of descriptors are similar requires to clarify what elements should be compared, and how the comparison should be enforced. We have tried two very different approaches where elements to compare were either the descriptors themselves, or sequence models learned from the values of the descriptors. Directly comparing sequences of descriptors is done using the traditional Dynamic Time Warping approach. It is in fact an *a posteriori* alignment of the sequences to compare. Here, the similarity of sequences is directly related to the similarity of the descriptions. We also compared sequence models, where each sequence is modeled using a Support Vector Machine approach used in regression (and not in classification, as usually done). Each model is somehow a translation of the temporal behavior of its corresponding sequence. Overall, we have shown that relying on models (instead of relying on descriptors) provides a better robustness to severe modifications of sequences, like temporal distortions for example. These results were obtained using a sequence collection made of real audio data broadcast on radio. Up to know, we used cross-similarity estimation based metrics to compare models as direct comparison between models is impossible. We are planning to investigate more distance-compliant metrics that would have better indexing properties.

These initial results suggest to push forward the investigations. We will look on ways to connect the models to multidimensional indexing techniques, as to handle scale.

6.1.3.6. Browsing Personal Image Collections

Participants: Laurent Amsaleg, Kári Harðarson.

Before entering into the details, it is worth saying that Kári Harðarson is again visiting TEXMEX for 4 months, starting September.

Our aim for the year 2008 was to write a prototype of a photo browser. A lot of time was spent finding the right tools for this prototype. The browser is now up and running in a basic version. It is written in the interpreted Python language, using the Qt library to generate the Graphical User Interface. So far, it uses a simplistic database scheme, sophisticated enough to provide the appropriate abstractions we need over a photo browsing mechanism.

We have worked quite extensively on the display part in order to display pictures in a very customizable way, and also such that classical problems of occlusions, image stacking get alleviated. We have also started putting together thoughts for facilitating the annotation of pictures to make great use of semantics.

We have also started working on more fundamental issues such as designing a complete data model for supporting pictures and their attributes. Currently, the model is half-way between OLAP principles and the Logical Information Systems invented by our colleague Sébastien Ferré (LIS team, IRISA).

6.2. New techniques for language processing and applications

6.2.1. NLP for Document Description

6.2.1.1. Indexing Biomedical Documents

Participant: Vincent Claveau.

This work is done in collaboration with Aurélie Névéol, National Institute of Health, USA.

In the biomedical domain, most of the publications are manually indexed with a reference set of words (the MeSH - Medical Subject Headings) by human indexers from the NIH for the well-known repository Medline. Every publication is given MeSH terms and each term can be detailed by some modifiers corresponding to the content of the publication. Some tools exist to help assign automatically MeSH terms to article, but nothing for the modifiers. With A. Névéol, we have shown that some of these modifiers could be simply deduced from all the MeSH terms used to describe an article [38]. We used a machine learning approach, namely inductive logic programming, and adapted it to infer rules explaining the complex connections between the modifiers and the terms [46]. The whole process is unsupervised. These ilp-based rules are now being integrated in the tools used by the indexers for Medline [15].

6.2.2. Oral and Textual Information Retrieval

Keywords: *Information Retrieval*, *Machine Learning*, *Natural Language Processing*.

6.2.2.1. Information Retrieval in the TV context

Keywords: *Automatic Speech Recognition*, *Information Retrieval*, *Phoneme-level Transcriptions*, *TV Data*.

Participant: Fabienne Moreau.

The work on this topic is done in close collaboration with Guillaume Gravier from the METISS project-team.

The main focus of our research is to use IR (Information Retrieval) systems on TV data. Classical approach uses ASR (Automatic Speech Recognition) system to transcribe the speech contained in TV streams, combined with a standard text-based indexing in order to retrieve information from these transcriptions. But transcriptions provided by ASR often contain a considerable number of word recognition errors, especially in the TV context where data is very noisy, and the transcription quality greatly varies depending on TV programs. In that context, we studied: 1. various approaches to account for transcription errors (confidence measures, confusion networks, phonetic distance...): we verified that the performances of our IR system applied on transcriptions with high error rates are improved using, *e.g.*, confusion networks. 2. phonetic approaches to deal with the issue of the out-of-vocabulary words (*i.e.* absent from the ASR system vocabulary), which occurs frequently in the TV context.

In the framework of the Star Challenge (cf. Section 8.2.4), we extended phonetic approaches to deal with multiple language collection of documents. To compensate for the imprecision of the document phonetic transcriptions, we proposed to expand queries (represented by phoneme sequences) with additional phonetic knowledge. Information used to enrich original queries is obtained from phone confusion matrices that identify, for a given language, groups of phonemes that are more likely to be confused with each other.

6.2.2.2. Cross-Language Information Retrieval in the Biomedical Domain

Participant: Vincent Claveau.

In the biomedical domain, the international research framework makes knowledge resources such as multilingual terminologies and thesauri essential to carry out many researches. Within this context, we developed an original method to translate biomedical terms from one language to another, relying on the proximity of such terms in many languages (for instance, consider the French-English examples: ophtalmorragie/ophthalmorrhagia, ophtalmoplastie/ophthalmoplasty, leucorragie/leukorrhagia). The main idea of our work is to automatically learn these regularities with well suited machine-learning techniques, and then to use them to translate new or unknown biomedical terms. Following our work of 2007, our approach is based on rewriting rules learning and on language modeling. It allows us to generate translations of terms between various languages (English, French, Italian, Portuguese, Russian...) with very good results [25]. The complexity of our approach and its usefulness in a cross-language information retrieval task have also been detailed [48].

6.2.2.3. Graded-Inclusion-Based Information Retrieval Systems

Keywords: *Fuzzy Division*, *Fuzzy Logic*, *Graded Inclusion*, *IR Models*.

Participants: Vincent Claveau, Laurent Ughetto.

Our work on this topic is done in close collaboration with Olivier Pivert and Patrick Bosc from the PILGRIM *team of IRISA Lannion.*

Databases (DB) querying mechanisms, and more particularly the division of relations was at the origin of the Boolean model for Information Retrieval Systems (IRSs). This model has rapidly shown its limitations and is no longer used in IR. Among the reasons, the Boolean approach does not allow to represent and use the relative importance of terms indexing the documents or representing the queries. However, this notion of importance can be captured by the division of fuzzy relations. This division, modeled by fuzzy implications, corresponds to graded inclusions. Theoretical work conducted by the PILGRIM project-team have shown the interest of this operator in IR.

Our first work was to investigate the use of graded inclusions to model the information retrieval process. In this framework, documents and queries are represented by fuzzy sets, which are paired with operations like fuzzy implications and T-norms. Through different experiments, we have shown that only some among the wide range of fuzzy operations are relevant for information retrieval. When appropriate settings are chosen, it is possible to mimic classical systems, thus yielding results rivaling those of state-of-the-art systems. These positive results have validated the proposed approach, while negative ones have given some insights on the properties needed by such a model.

From these encouraging results, perspectives have be derived, and will be addressed next year. For instance, this graded inclusion-based model gives new and theoretically grounded ways for a user to easily weight his query terms, to include negative information in his queries, or to expand them with related terms.

6.3. New processing tools for audiovisual documents

6.3.1. TV Stream Structuring

6.3.1.1. TV stream macro-segmentation

Participants: Patrick Gros, Gaël Manson.

This work is done in the frame of the CIFRE thesis of Gael Manson at Orange Labs (former France-Telecom R&D).

TV stream macro-segmentation basically aims at precisely determining the start and the end of each of the broadcasted programs and inter-programs. A first step in automatic TV macro-segmentation is segmenting TV stream in programs and inter-programs segments. The most promising automatic approach for that consists in detecting inter-programs as near-identical repeated sequences. However, the resulting repeated sequences can only detect sequences that repeat in the stream. Some of these sequences are actually inter-programs and others belong to long programs. Moreover, some inter-programs do not repeat.

It is therefore necessary to classify the segments resulting from the repeated sequences detection phase (the occurrences of repeated sequences and the rest of the stream) into program and inter-program segments. Our solution for that is based on Inductive Logic Programming. In addition to intrinsic features of each segment (e.g. duration, number of repetitions), our technique makes use of the relational and contextual information of the segments in the stream (e.g. the class of following segments, the class of other occurrences of repetitions).

On the other hand, we have performed a deeper study on the repetition property of inter-programs. This study shows that inter-programs repeat sufficiently to perform a macro-segmentation and that a simple rule-based technique using detected repeated sequences allows to accurately macro-segment a TV stream.

The following step we are working on is how to go further in the classification process and its application for TV stream macro-segmentation. In this way, each segment could be classified as a commercial, a program (or a part of), a sponsoring, a jingle, a trailer or a opening/closing credit.

6.3.1.2. Repetition Detection-based TV Structuring

Participants: Patrick Gros, Zein Al Abidin Ibrahim, Emmanuelle Martienne.

This work is done under the Semim@ges project in collaboration with the CAIRN team.

In the technique explained above, the detection of the repeated segments is mainly used for segmentation purposes. The classification of the resulting segments is based on local information (time and day of diffusion, duration, class of the neighboring segments...) In the frame of the Semim@ges project, we try to develop another approach where the detection of the repeated segments is used for the classification process itself. The basic idea is that the various classes of sequences (advertisements, self promotion, trailers...) follow different repetition patterns, and thus can be recognized according to their own pattern.

The first step of the method is to detect repeated segments. This can be achieved efficiently with hashing techniques and a filtering method to accelerate the research. It takes less than three minutes to detect the repeated sequences in a TV stream composed of three consecutive weeks. An analogous work is done by the CAIRN team of IRISA with a hardware acceleration. They rely on reprogrammable circuit of type FPGA to speed up the computation.

The second step of this work is to cluster the sequences basing on their pattern of repetitions. Based on a symbolic description of the repeated sequence, we use a unsupervised hierarchical clustering method to cluster the repetitions. The clusters are then annotated manually. Aside this numerical method, we also use a symbolic learning technique, Inductive Logic Programming (ILP) in order to compare the possibilities of both approaches. Clustering techniques imply to used fixed length description of the sequences while ILP allow to have variable length descriptions.

Finally, the last step consists in aligning this stream with an EPG using the DTW algorithm.

6.3.1.3. Semantic TV Program Verification

Participants: Camille Guinaudeau, Gwénolé Lecorvé, Pascale Sébillot.

Our work on this topic is done in close collaboration with Guillaume Gravier from the METISS *project-team.*

Over the previous years, TEXMEX has developed an approach for the automatic labeling of program segments from the electronic program guide (EPG) [77]. In order to validate the labeling from a semantic point of view, we associate the phonetic or textual transcription of the soundtrack with descriptions extracted from an online TV guide for each segment. This association, which, as opposed to the labeling step, relies on semantic content, is carried out with techniques inspired from information retrieval methods. Names obtained from the TV guide are then compared with the respective labels obtained from the EPG alignment. The phonetic and textual methods implemented allow to make a decision for 40 % of the segments and to decrease the labeling error rate by 3.5 % with respect to X. Naturel's method.

6.3.2. Program Structuring

6.3.2.1. Stochastic Models for Video Description

Keywords: *Dynamic Bayesian Networks*, *Hidden Markov Models*, *Image-Sound Interaction*, *Segment models*, *Video Structuring*.

Participants: Siwar Baghdadi, Patrick Gros.

Our work on this topic is done in close collaboration with Guillaume Gravier from the METISS *project-team and Thomson as external partner.*

Bayesian Networks are an elegant and powerful semantic analysis tool. They combine an intuitive graphical representation with efficient algorithms for inference and learning. They also allow the representation in a comprehensive manner of the interaction between a system variable and the integrating of external knowledge. Unlike HMM and segment models, structures of Bayesian Networks are very flexible and can be learned from data. We explored the idea of using Bayesian Networks and their temporal extension Dynamic Bayesian Networks to do event detection in video streams. As a first application we have chosen commercial detection. We modeled the video stream by a Dynamic Bayesian Network. According to this model, the video stream is a sequence of observations (a set of multimodal features). Each observation is generated according to the state of the system (program or commercial). The model is fed with knowledge about the commercial segment duration. Detecting commercial segment is then a problem of inferring the optimal sequence of hidden nodes with the convenient duration. Structure learning allowed us to learn the optimal interaction between variables. Future work involves the extension of our model to do event detection in soccer games. The challenge of this part is to take into consideration all kinds of feature interactions (spatial, temporal at short or long term).

6.3.3. Using Speech to Describe and Structure Video

Keywords: *Natural Language Processing*, *Semantics*, *Speech Recognition*, *Video Structuring*.

Participants: Camille Guinaudeau, Stéphane Huet, Gwénolé Lecorvé, Pascale Sébillot.

Our work on this topic is done in close collaboration with Guillaume Gravier and Pierre Cauchy from the METISS *project-team.*

Speech can be used to structure and index large collections of spoken documents (videos, audio streams...) based on semantics. This is typically achieved by first transforming speech into text using automatic speech recognition (ASR), before applying natural language processing (NLP) techniques on the transcriptions. Our research focuses firstly on the adaptation of NLP methods designed for regular texts to account for the specificities of automatic transcriptions. In particular, we also investigate a deeper integration between ASR and NLP, *i.e.* between the transcription phase and the semantic analysis phase. Finally, we study the effective use of semantic analysis for video structuring.

In 2008, we have considered topic segmentation of streams and segment characterization.

We improved our former extension of the text-based topic segmentation method of Utiyama and Isahara [89] to take into account additional knowledge such as semantic relations between words, discourse markers (like `"and now, thank you"`), and acoustic cues [42]. Results obtained on radio broadcast news make it possible to apply the method to large scale TV streams, eventually in conjunction with image-based features, as considered in C. Guinaudeau's Ph. D. thesis.

In the framework of G. Lecorvé's Ph. D. thesis, we investigated efficient methods for the extraction of keywords to characterize thematic segments, in order to improve the language model of the ASR system using related texts retrieved on the Internet. Experiments reported in [30] have shown that topic adaptation is more effective when included in the early recognition stages. Thus, we focused on keyword extraction at the very beginning of the transcription using confusion networks rather than a single sentence.

This year has also been dedicated to a major refactoring of the ASR system, in collaboration with the METISS project-team, using a larger amount of training data. In particular, we have redesigned the language model (LM) of the ASR system in order to enable vocabulary adaptation on top of LM adaptation.

Keyword extraction and retrieval of topic-related documents on the Internet offers many opportunities for video structuring and browsing. In particular, it enables linking segments of a video with external related documents: for example, articles of online newspapers can be linked to broadcast news segments or film summaries and reviews to the corresponding part of the TV stream. Enriching video streams with external documents is considered in the Ph. D. thesis of Camille Guinaudeau which started in Oct. 2008 (supervised by P. Sébillot and G. Gravier).

7. Contracts and Grants with Industry

7.1. Contracts, Initiatives and Participation to Networks of Technological Research

7.1.1. Pôle de Compétitivité Image et Réseaux

Participant: Patrick Gros.

The French government organized in 2005 competitiveness clusters (*pôles de compétitivité*) in France to strengthen ties in a given region between industries (big and small companies), research labs (both public and private ones) and teaching institutions (universities and schools of engineering). We are part, through our participation to the two projects Semim@ges and ICOS-HD, to the pole called "Images and networks" whose main actors are Thomson and Orange Labs and which is located in Brittany and Pays de la Loire; P. Gros is deputy administrator of the cluster.

7.1.2. Contract with Thomson

Participants: Siwar Baghdadi, Patrick Gros.

Duration: 36 months, starting in December 2005.

S. Baghdadi's Ph.D. thesis is supported by a CIFRE grant in the framework of a contract between Thomson and TEXMEX.

7.1.3. Contract with Orange Labs

Participants: Patrick Gros, Gaël Manson.

Duration: 36 months, starting in November 2006.

G. Manson's Ph.D. thesis is supported by a CIFRE grant in the framework of a contract between Orange Labs and TEXMEX.

7.1.4. Quaero

Participants: Laurent Amsaleg, Mathieu Ben, Vincent Claveau, Patrick Gros, Camille Guinaudeau, Ewa Kijak, Stacy Payne, Pascale Sébillot.

Duration: 5 years, starting in May 2008. Prime: Thomson.

Quaero is a large research and applicative program in the field of multimedia description (ranging from text to speech and video) and search engines. It groups 5 application projects, a joint Core Technology Cluster developing and providing advanced technologies to the application projects and a Corpus project in charge of providing the necessary data to develop and evaluate the technologies. 6 project-teams of INRIA participate to this program.

TEXMEX is involved on three main topics: Multidimensional indexing with application to multimedia data identification, Natural Language processing and TV structuring and analysis. Our major collaborators in the program are INA, Thomson, Exalead and LTU.

7.2. European Initiatives

7.2.1. European Network of Excellence MUSCLE: Multimedia Understanding through Semantics, Computation, and Learning

Keywords: *Images*, *Multimedia*, *Natural Language Processing*, *Video*.

Participants: Laurent Amsaleg, Ewa Kijak, Gwénolé Lecorvé, Patrick Gros, Pascale Sébillot.

Duration: 4 years, starting in April 2004. 42 partners. Prime: ERCIM, scientific coordinator: Nozha Boujemaa, INRIA.

This project aims at developing the collaboration in the domain of automatic multimedia document analysis, in particular to be able to handle and exploit their meaning. Thus, it is concerned by all content-based analysis tools available for every media (text, sound and speech, image and video), but also by the techniques which allow us to combine the information extracted from each media, and by the common techniques needed to handle such data (optimization, classification, intensive computation).

P. Gros co-edited a book on multimodal processing and interaction [57] with P. Maragos from the National Technical University of Athens, and A. Potamianos from Technical University of Chania. The book was published by Springer in 2008.

7.3. Start-up Creation

7.3.1. Powedia

Participant: Arnaud Dupuis.

Powedia is a start-up interface for the technology transfer of the TEXMEX *Team.*

The Powedia start-up project is the result of research and development activities led for more than three years within TEXMEX. It proposes to commercialize a new technology for dynamic video editing. Particularly, it aims at offering new video services and to make more profitable audiovisual contents broadcasted on web sites.

The technological innovation is the capacity to realize dynamic video editing, adapted to the profile or the geographical situation of Internet users. The system is very flexible, and not intrusive. It adapts by itself to existing web sites and proposes a legal and free solution to share audiovisual contents for Internet users.

Thus, this software solution allows media to propose new broadcasting services such as the editing or legal sharing, while reducing operating costs and multiplying the sources of profits, thanks to the addition of targeted advertising campaigns. It offers the advantage to combine the current uses of content sharing for Internet users, with the need to grant legal incomes for the authors.

Powedia integrated the Emergys incubator in April 2008 and was an award winner of the 10th *concours National d'aide à la création d'entreprises de technologies innovantes, catégorie "en émergence"* organized by the French ministry of Education and Research.

Powedia will commercialize his software solutions on the 1st quarter of 2009.

8. Other Grants and Activities

8.1. National Initiatives

8.1.1. ACI masses de données DEMI-TON: Multimodal Description for Automatic Structuring of TV Streams

Participants: Sébastien Campion, Arnaud Dupuis, Patrick Gros, Stéphane Huet, Gwénolé Lecorvé, Pascale Sébillot.

Duration: 3 years, starting in April 2005. Partners: INA, METISS *project-team.*

This project concerns the development of new techniques to index large collections of TV programs. INA records and index more than 50 channels 24 hours a day. As the number of available archivists did not increase as fast as the number of channels to index, they have to rely on more automatic processes. The first need is to verify that the programs in the stream correspond effectively to what was announced in the TV program guide and to synchronize the stream with this program guide. In a second stage, some programs like news reports have to be indexed at a finer grain: we want to describe each topic that was tackled by the program and which, of course, could not be announced in the program guide.

After two years mainly dedicated to software development, our video analysis platform was mainly upgraded on the hardware side this year. A NAS server of 100 terabytes of memory was installed with several side servers to record and distribute data. These machines are connected to the grid of IRISA (more than 400 cores) to obtain the necessary computing power.

The work initiated within DEMI-TON on TV stream structuring through the thesis of X. Naturel and G. Lecorvé was continued in the Semim@ages project and will go on in the frame of the QUAERO program. DEMI-TON was evaluated during the summer and considered as very successful by the evaluators.

8.1.2. ANR project ICOS-HD

Participants: Ariane Herbulot, Ewa Kijak, Joaquin Zepeda.

Duration: 3 years, starting in January 2007. Partners: University of Bordeaux 1, CNRS-I3S.

This project concerns scalable indexing and compression for high definition video content management. Recent solutions for achieving high-quality compression of images/video resulting in scalable bit streams. The objective of the project is to propose new solutions of scalable description to facilitate editing, manipulation and access of HD contents via heterogeneous infrastructures. TEXMEX project-team is involved on the study of new signal representation amenable to both compression and image description, as well as descriptor adaptation for image retrieval in large databases.

8.1.3. ANR project Semim@ges

Participants: Laurent Amsaleg, Sébastien Campion, Patrick Gros, Emmanuelle Martienne.

Duration: 27 months, starting in January 2007. Partners: Orange Labs, TDF, Kersonic, Telisma, CAIRN team.

The project is devoted to TV data exploitation and repurposing. Two main applications will be considered: TV news analysis, and TV streams structuring. TEXMEX project-team will mainly be involved in the second one. The aim of our work is to structure automatically long TV streams in more usable units like programs or non-program sequences, exactly like it was done in Xavier Naturel's thesis. But we would like to achieve this same goal completely automatically this time, by removing most of the manual annotation required and by relying on non-supervised classification techniques.

The first step of this work consists in detecting all replicated images in the stream. This will be done by a massively parallel approach in collaboration with the CAIRN team.

8.1.4. ARC INRIA RAPSODIS: Syntactic and Semantic Information-Based Automated Speech Recognition

Participants: Camille Guinaudeau, Stéphane Huet, Gwénolé Lecorvé, Pascale Sébillot.

Duration: 2 years, starting in February 2008. Partners: METISS, PAROLE, TALARIS *project-teams, CEA-LIST/LIC2M.*

This project aims at improving automatic speech recognition (ASR) by integrating linguistic information. Based on former work by S. Huet concerning the incorporation of morpho-syntactic knowledge in a post-processing stage of the transcription, we experiment, together with our partners, the deep insertion of automatically obtained semantic relations (especially paradigmatic ones) and syntactic knowledge within an ASR system.

In 2008, work has been mostly dedicated to the study of possible integration modes—reordering of n-best hypothesis lists is currently privileged—, to investigations about the impact of transcription errors on syntactic parsing—a correlation between the length of chunks in hypothesis and errors has been established—, to the acquisition of semantic relations from the Web, and to a major refactoring of the METISS project-team's ASR system, using a larger amount of training data, in order to enable effective integration of linguistic information.

8.2. International Collaborations

8.2.1. Collaboration with Reykjavík University, Iceland

Keywords: *Approximate Search Schemes*, *Content-Based Image Retrieval Systems*, *Curse of Dimensionality*, *Local Descriptors*, *Random Projections*, *User Interface*.

Participant: Laurent Amsaleg.

This collaboration is done in the context of the INRIA Associate Teams program. This program links two research teams (one INRIA, one foreign) willing to cross-leverage their respective excellence and their complementarity. Björn Þór Jónsson (Associate Professor) leads the team of researchers involved in Iceland.

Image databases, and content-based image retrieval systems in particular, have become increasingly important in many applications areas. While extremely effective (they return high quality results), these systems are very inefficient (they answer very slowly) due to their complexity, to the curse of dimensionality problems and to the scale at which they have to run when dealing with collections of realistic sizes.

Originally, the goal of this project was to research and develop new database support that integrates efficiency and effectiveness for modern, large-scale, computer-vision related applications and problems. We found, however, that the inefficiency observation applies not only to images, but to most multimedia documents as soon as the retrieval is based on the similarity of low-level descriptions.

Therefore, the Eff2 associate team has widen its scope and tackles efficient database support for more general large-scale multimedia applications and problems that mostly deal with high-dimensional low-level features. We now investigate additional issues such as efficient query execution on large collections of sequences of low-level descriptions (e.g., audio or video collections). We have also initiated another thread of research by investigating the browsing of personal image collections. Going to sequences and addressing browsing issues (and not solely searching) is a new development in our cooperation.

8.2.2. Collaboration with Croatia and Slovenia

Participant: Annie Morin.

Medical School, University of Zagreb, department of Electronics, Microelectronics, Computer and Intelligent systems, University of Zagreb, Zagreb, Croatia; Faculty of Computer and Information Science, University of Ljubljana, Slovenia; ERIC lab., University of Lyon2

A. Morin got two Egide contracts with Slovenia (Proteus) and Croatia (Cogito) for 2007 and 2008 on knowledge discovery and visualization for textual data. In Slovenia, we work with Blaz Zupan and Janez Demsar from faculty of Computer and Information Science, University of Ljubljana and in Croatia with Bojana Dalbelo Bašić from faculty of Electrical Engineering and Computing, university of Zagreb. The French laboratory ERIC (university of Lyon 2) is the other French partner.

The concerned research teams have different expertise on the same subject: machine learning for the Slovenian and Croatian teams, statistics for the French teams and common abilities such as development of open source data mining software and visualization tools. They have been in touch since a first meeting in 2004 on intelligent data mining. We plan to implement a new visualization system for textual data. Proposed collaboration includes sharing of a number of Ph.D. students.

Artur Šilić, a student from the University of Zagreb, has got an internship and has spent 3 months in our team. A. Šilić started to work on temporal changes in text streams.

8.2.3. Collaboration with Nagoya University and National Institute of Informatics, Japan

Participants: Laurent Amsaleg, Patrick Gros, Fabienne Moreau, Annie Morin, Pascale Sébillot.

In 2008, we started a new associated team with both University of Nagoya (Ichiro Ide, Associate Professor) and the National Institute of Informatics in Tokyo (Shin'ichi Satoh, Professor). The work done in the frame of this collaboration concerns topic threading, TV structuring. This year, we had a joint participation to the Star Challenge. (see next-subsection and Section 8.3.4 for further details).

8.2.4. Participation to The Star Challenge evaluation campaign

Participant: Fabienne Moreau.

This work was done in collaboration with Guillaume Gravier from the from the METISS *project-team.*

In the framework of the INRIA associate team e-TV, we participated in The Star Challenge, an international competition on information retrieval tasks in multimedia contents, as part of an international joint venture with the National Institute of Informatics (NII, Tokyo) and the METISS project-team. Our joint team passed the qualifying race and took part of the grand final held in Singapore, where it ranked second out of the five finalist competitors.

TEXMEX and METISS were involved in audio retrieval tasks, such as the "Query by IPA" one which consisted in retrieving shots containing a given string of phonemes in a multilingual collection of video shots. Details of the research activities undertaken in the framework of our participation to The Star Challenge are given in Section 6.2.2.

8.3. Visits of foreign researchers, Invitations to foreign labs

8.3.1. Visit of Members of University of Ljubljana

Participant: Annie Morin.

Lan Umek, a Ph.D student from the university of Ljubljana, came in January 2007 and spent 5 weeks in TEXMEX project working on the comparison of indicators in decision trees.

8.3.2. Visit of Members of University of Zagreb

Participant: Annie Morin.

Bojana Dalbelo Bašić came twice in Rennes in April 2008 to work on a paper we submitted to IDA journal and in september 2008 to prepare a project for associate team. Saša Petrovic was here also during a week in April to work on the paper. Besides, Artur Šilić spent 3 months from April to July 2008. In 2007, Artur Šilić spent a week in Lyon (laboratory involved in our Cogito program) working on n-grams and morphological normalization in text classification. The work was presented at the Portuguese conference of Artificial Intelligence in December 2007 and has been published in LNCS in 2008. For this paper, A. Šilić has won the Croatian Award "Science" given by the National Science Foundation of Croatia to an undergraduate student who manages to publish a paper in a publication with the highest citation index. Currently Artur Šilić is working on text streams topic.

8.3.3. Visit to the University of Zagreb

Participant: Annie Morin.

A. Morin spent 1 week at the university of Zagreb in June 2008 and attended the ITI conference in Cavtat just after this stay.

8.3.4. Visit to the University of Nagoya and NII Tokyo

Participants: Patrick Gros, Pascale Sébillot.

Partners: Ichiro Ide, Associate professor at the University of Nagoya, and visiting associate professor at NII (National institute of informatics) Tokyo, Japan, and Shin'ichi Satoh, Professor at NII Tokyo, Japan.

In the framework of the INRIA Associate team between TEXMEX, the University of Nagoya and NII Tokyo, Patrick Gros visited both institutions in February 2008, in order to study the opportunity of a common project in the frame of the joint ANR-JST call for proposal. Unfortunately, our proposal was finally not accepted.

Guillaume Gravier (METISS project-team) and Pascale Sébillot spent 2 weeks in Japan (both at Nagoya and Tokyo) in May 2008. During this period, they worked on various aspects of automatic structuring of TV streams: creation of a temporally-similar broadcast news corpus in Japanese and French to carry on similar investigations, relevance of automatic speech recognition in multimedia information retrieval...

8.3.5. Visit from Members of University of Nagoya and NII Tokyo

Participants: Camille Guinaudeau, Gwénolé Lecorvé, Pascale Sébillot.

Partner: Ichiro Ide, Associate professor at the University of Nagoya, and visiting associate professor at NII (National institute of informatics) Tokyo, Japan.

In the framework of the INRIA Associate team between TEXMEX, the University of Nagoya and NII Tokyo, Ichiro Ide spent 2 days in the team in December 2008. These two days were dedicated to discussions and common work about the exploitation of the temporally-similar broadcast news corpus in Japanese and French the elaboration of which has begun in July. The aim of this project is to study topic tracking on both streams and investigate differences between the channels.

9. Dissemination

9.1. Conference, Workshop and Seminar Organization

- L. Amsaleg was the general chair of the workshops co-located with the 11th International Conference on Extending Database Technology (EDBT 2008), Nantes, France, March 25–29, 2008.
- L. Amsaleg and V. Claveau organized and edited the proceedings of the 5th edition of the "COnférence en Recherche d'Information et Applications (CORIA 2008)", Trégastel, France, March 12–14, 2008.

9.2. Involvement with the Scientific Community

- L. Amsaleg:
  - was a program committee member of 2008 ACM International Conference on Multimedia, Vancouver, BC, Canada;
  - was a program committee member of Bases de données avancées, BDA 2008, Guilherand-Granges, France;
  - was a program committee member of the 9th International Workshop on Multimedia Data Mining, associated with the Fourteenth ACM SIGKDD International Conference on Knowledge Discovery and Data Mining (KDD08), Las Vegas, NV, USA;
  - was on the reading committee of the International Journal on Very Large Data Bases, 2008;
  - was a program committee member of the Fourteenth International Multimedia Modeling Conference, Kyoto, Japan;
  - was a program committee member of the 17th International World Wide Web Conference, Beijing, China;
  - is the 2008 ACM SIGMOD Workshops Coordinator.
- V. Claveau:
  - was a program committee member of TALN'08 (15^{e} conférence nationale Traitement automatique des langues naturelles), Toulouse, France, June 2008;

  - was a program committee member of KES'08 (12th International Conference on Knowledge-Based and Intelligent Information & Engineering Systems), Zagreb, Croatia, September 2008;
  - was a program committee member of FODOP'08 (Atelier FOuille des Données d'OPinions), Fontainebleau, France, May 2008.

- P. Gros:
  - is a deputy administrator of the "Pôle de compétitivité Images et Réseaux".
  - was a program committee member of the fifth COnference on Information Retrieval and Application CORIA'08, Trégastel, March 2008;
  - was a program committee member of the 4th International Symposium on Image/Video Communications over fixed and mobile networks, Bilbao, July 2008;
  - was a program committee member of the 14th Multimedia Modeling Conference, Kyoto, January 2008;
  - was a program committee member of the 16th "congrès francophone AFRIF-AFIA Reconnaissance des Formes et Intelligence Artificielle", Amiens, January 2008;
  - was a program committee member of the GRETSI conference, Troyes, June 2007.
  - was a program committee member of the International conference on signal processing and multimedia applications SIGMAP'08, Milan, July 2008.

- F. Moreau:
  - was a program committee co-chair of RJCRI'08 (Rencontres Jeunes Chercheurs en Recherche d'Informations), Trégastel, France, March 2008.

- A. Morin:
  - is a program committee member of ITI 2008 (Information technology interfaces);
  - is vice-president of the CNU (National Council of the University) in the computer science section.

- F. Poulet:
  - was a program committee member of AusDM'08, Australasian Data Mining Conference, Stamford Grand, Glenelg, Adelaide, November 2008;
  - was a program committee member of EGC'08, Extraction et Gestion de Connaissances, Sophia-Antipolis, January 2008;
  - was co-organizer of the 6th workshop Visualisation et Extraction de Connaissances (EGC'08), Sophia-Antipolis, January 2008;
  - was a reviewing committee member for Neurocomputing journal;
  - was a reviewing committee member for the International Journal of Computational Methods;
  - was a co-editor of a special issue of Revue d'Intelligence Artificielle (vol.22, n°3-4) on Visualization and Knowledge Discovery;

- P. Sébillot:
  - was a program committee member of RFIA 2008 (16e congrès francophone AFRIF-AFIA Reconnaissance des formes et intelligence artificielle), Amiens, France, January 2008;
  - was a program committee member of JADT2008 (9e journées internationales d'analyse statistique des données textuelles, Lyon, France, March 2008;

- was a scientific committee member of LREC 2008 (6th Language Resources and Evaluation Conference), Marrakech, Morocco, May 2008;
- was a program committee member of TALN 2008 (15e conférence nationale Traitement automatique des langues naturelles), Avignon, France, June 2008;
- was a review committee member of KES2008 (12th International Conference on Knowledge-Based and Intelligent Information and Engineering Systems, Zagreb, Croatia, September 2008.

- P. Tirilly:
  - was a program committee member of RJCRI'08 (Rencontres Jeunes Chercheurs en Recherche d'Informations), Trégastel, France, March 2008.
- L. Ughetto:
  - was a program committee member of the Rencontres Francophones sur la Logique Floue et ses Applications (LFA'08), Lens, France, October 2008.

9.3. Teaching Activities

- L. Amsaleg and F. Poulet: Managing Large Collections of Digital Data. Master by research in computer science (2nd year), University of Rennes 1.
- L. Amsaleg: Advanced Databases, ENSAI.
- L. Amsaleg: Multidimensional Indexing Techniques, University of Reykjavik, Master.
- P. Gros coordinates the track "From Data to Knowledge: Machine Learning, Modeling and Indexing Multimedia Contents and Symbolic Data" of the Master by research in computer science (2nd year), University of Rennes 1.
- E. Kijak: Analysis of audiovisual documents and flows for indexing, Master by research in computer science (2nd year), University of Rennes 1.
- E. Kijak and P. Tirilly: Digital Documents Indexing and Retrieval, Professional Master in Computer Science, 2nd year, IFSIC, University of Rennes 1.
- F. Poulet: Introduction to Data Mining. Professionnal Master in Computer Science, 2nd year, IFSIC, University of Rennes 1.
- P. Sébillot is in charge of the Master by research in computer science (2nd year), University of Rennes 1.
- P. Sébillot, V. Claveau and P. Tirilly: Advanced Databases and Modern Information Systems, 5th year, Computer Science, INSA Rennes.
- V. Claveau: Symbolic Sequential Data, Master by research in computer science (2nd year), University of Rennes 1.

9.4. Participation to Seminars, Workshops, Invited Conferences

- L. Amsaleg gave an invited talk in the context of the Wisdom seminars federating the three database research teams located in downtown Paris.
- P. Gros was invited to give a talk to ARMT (Autorité de régulation des mesures techniques, an official body in charge of proposition technical solutions to reduce piracy of pictures, movies and music on Internet) to present content-based retrieval and watermarking in the context of copy detection.
- L. Amsaleg and P. Gros were invited to give several talks at Reykjavík university about image description and multidimensional indexing in September 2008.

- E. Kijak gave an invited talk at the Reykjavík University about multimedia indexing, in December 2008.
- P. Sébillot gave an invited talk at the University of Nagoya about automatic speech recognition and natural language processing, May 2008.
- P. Sébillot gave an invited talk at the National institute of informatics (NII) Tokyo about information retrieval and natural language processing, May 2008.

9.5. Popular work, Press articles

- The TEXMEX team represented INRIA at the “fête de la science” event in Rennes in November 2008. This 3-day event aims at encouraging interactions between scientific labs and the public, through presentations, scientific games, demos...

10. Bibliography

Major publications by the team in recent years

[1] L. AMSALEG, P. GROS. *Content-based Retrieval Using Local Descriptors: Problems and Issues from a Database Perspective*, in "Pattern Analysis and Applications", vol. 2001, n^o 4, March 2001, p. 108-124.

[2] J. ANDRÉ, A. MORIN, H. RICHY. *Comparison of Literary Texts using Biological Sequence Comparison and Structured Documents Capabilities*, in "Proceedings of the International Conference on Computational Linguistics, Speech and Document Processing, ICCLSDP'98, Calcutta, India", February 1998.

[3] V. CLAVEAU, P. SÉBILLOT, C. FABRE, P. BOUILLON. *Learning Semantic Lexicons from a Part-of-Speech and Semantically Tagged Corpus Using Inductive Logic Programming*, in "Journal of Machine Learning Research, special issue on Inductive Logic Programming", vol. 4, August 2003, p. 493–525.

[4] M. DELAKIS, G. GRAVIER, P. GROS. *Multimodal Segmental-Based Modeling of Tennis Video Broadcasts*, in "Proceedings of IEEE International Conference on Multimedia and Expo, Amsterdam, Netherlands", July 2005, http://www.irisa.fr/texmex/publications/versionElect/2005/icme2005_manolis.pdf.

[5] E. KIJAK, G. GRAVIER, L. OISEL, P. GROS. *Audiovisual integration for sport broadcast structuring*, in "Multimedia Tools and Applications", vol. 30, 2006, p. 289-312, http://www.springerlink.com/content/24h61433843r474l/.

[6] H. LEJSEK, F.-H. ÁSMUNDSSON, B.-Þ. JÓNSSON, L. AMSALEG. *Scalability of Local Image Descriptors: A Comparative Study*, in "Proceedings of the 14th ACM International Conference on Multimedia, Santa Barbara, CA, USA", October 2006, http://www.irisa.fr/texmex/publications/versionElect/2006/fp33a-lejsek.pdf.

[7] R. PRIAM, A. MORIN. *Visualisation des corpus textuels par treillis de multinomiales auto-organisées - Généralisation de l'analyse factorielle des correspondances*, in "Revue Extraction des Connaissances et Apprentissage (Actes EGC'02)", vol. 1, n^o 4, 2002, p. 407-412.

[8] M. ROSSIGNOL, P. SÉBILLOT. *Combining Statistical Data Analysis Techniques to Extract Topical Keyword Classes from Corpora*, in "IDA (Intelligent Data Analysis)", vol. 9, n^o 1, 2005, p. 105-127.

[9] F. TONNIN, P. GROS, C. GUILLEMOT. *Analysis of multiresolution representations for compression and local description of images*, in "Proceedings of the 8th International Conference on Visual Information and Information Systems, VISUAL'05, Amsterdam, The Netherlands", 2005.

Year Publications

Articles in International Peer-Reviewed Journal

[10] S.-A. BERRANI, G. MANSON, P. LECHAT. *A non-supervised approach for repeated sequence detection in TV broadcast streams*, in "Signal Processing: Image Communication, special issue on Semantic Analysis for Interactive Multimedia Services", vol. 23, n^o 7, August 2008, p. 525-537.

[11] M. DELAKIS, G. GRAVIER, P. GROS. *Audiovisual Integration with Segment Models for Tennis Video Parsing*, in "Computer Vision and Image Understanding", vol. 111, n^o 2, August 2008, p. 142-154.

[12] H. LEJSEK, F.-H. ÁSMUNDSSON, B.-Þ. JÓNSSON, L. AMSALEG. *NV-tree: An Efficient Disk-Based Index for Approximate Search in Very Large High-Dimensional Collections*, in "IEEE Transactions on Pattern Analysis and Machine Intelligence", vol. 99, n^o 1, May 2008, http://www2.computer.org/portal/web/csdl/doi/10.1109/TPAMI.2008.130.

[13] X. NATUREL, P. GROS. *Detecting Repeats for Video Structuring*, in "Multimedia Tools and Application", vol. 38, n^o 2, May 2008, p. 233-252, http://www.springerlink.com/content/f417v67462m89067/.

[14] P. NGUYEN-KHANG, A. MORIN, P. GROS. *CAViz, An interactive graphical tool for image mining*, in "Journal of Computing and Information Technology (CIT)", vol. 16, n^o 4, 2008, p. 111-118.

[15] A. NÉVÉOL, S. E. SHOOSHAN, V. CLAVEAU. *Automatic inference of indexing rules for MEDLINE*, in "BMC BioInformatics", vol. 9, Supp. 11, November 2008, http://www.biomedcentral.com/1471-2105/9/S11/S11.

[16] L. UGHETTO, W.-A. VOGLOZIN, N. MOUADDIB. *Database Querying with Personalized Vocabulary Using Data Summaries*, in "Fuzzy Sets and Systems", vol. 159, n^o 15, August 2008, p. 2030–2046.

Articles in National Peer-Reviewed Journal

[17] L. BOUDJELOUD ASSALA, F. POULET. *Algorithme interactif pour la sélection de dimensions en détection d'outliers*, in "Revue d'Intelligence Artificielle (RIA)", Numéro spécial sur la visualisation et l'extraction des connaissances, vol. 22, n^o 3-4, 2008, p. 401-420.

[18] T.-N. DO, J.-D. FEKETE, F. POULET. *Algorithmes rapides de boosting de SVM*, in "RNTI-E-11, Revue des Nouvelles Technologies de l'Information", vol. 1, 2008, p. 297-308.

[19] E. FANGSEU BADJIO, F. POULET. *Théorie de consensus appliquée au prétraitement des ensembles de données*, in "RIA, Revue d'Intelligence Artificielle", vol. 22, n^o 3-4, 2008, p. 383-400.

[20] P. NGUYEN-KHANG, T.-N. DO, F. POULET, A. MORIN. *Tree-View : Exploration interactive des arbres de décision*, in "RIA, Revue d'Intelligence Artificielle", Numéro spécial sur la visualisation et l'extraction des connaissances, vol. 22, n^o 3-4, 2008, p. 473-487.

[21] P. NGUYEN-KHANG, A. MORIN, P. GROS. *CAViz, exploration interactive des résultats de l'analyse factorielle des correspondances pour des images*, in "Revue d'Intelligence Artificielle (RIA)", Numéro spécial sur la visualisation et l'extraction des connaissances, vol. 22, n^{o} 3-4, 2008, p. 489-502.

Articles in Non Peer-Reviewed Journal

[22] L. AMSALEG, B.-Þ. JÓNSSON, V. ORIA. *Introduction to special issue on Computer Vision meets Databases*, in "Multimedia Tools Applications", vol. 38, n^{o} 2, June 2008, p. 185-186.

[23] L. AMSALEG, B.-Þ. JÓNSSON, V. ORIA. *Report from the 3rd International Workshop on Computer Vision Meets Databases CVDB'07*, in "SIGMOD Record", vol. 37, n^{o} 1, March 2008, p. 46-48.

International Peer-Reviewed Conference/Proceedings

[24] S. BAGHDADI, G. GRAVIER, C.-H. DEMARTY, P. GROS. *Structure learning in a Bayesian network-based video indexing framework*, in "Proceedings of the IEEE International Conference on Multimedia and Expo (ICME 2008), Hanover, Germany", June 2008, p. 677-680.

[25] V. CLAVEAU. *Automatic Translation of Biomedical Terms by Supervised Machine Learning*, in "Proceedings of the 6th international conference on Language Resources and Evaluation (LREC'08), Marrakech, Morocco", May 2008, http://www.irisa.fr/texmex/people/claveau/publis/Claveau-LREC08.pdf.

[26] T.-N. DO, V.-H. NGUYEN, F. POULET. *A Fast Parallel SVM Algorithm for Massive Classification Tasks*, in "Proceedings of Modelling, Computation and Optimization in Information Systems and Management Sciences (MCO'08), Metz, France", October 2008, p. 419-428.

[27] T.-N. DO, V.-H. NGUYEN, F. POULET. *Speed up SVM Algorithms for Massive Classification Tasks*, in "Proceedings of Advanced Data Mining and Applications (ADMA'08), Chengdu, China", C. TANG, C. X. LING, X. ZHOU, N. CERCONE, X. LI (editors), Lecture Notes in Computer Science, vol. 5139, Springer-Verlag, October 2008, p. 147-157.

[28] S. HUET, G. GRAVIER, P. SÉBILLOT. *Morphosyntactic Resources for Automatic Speech Recognition*, in "Proceedings of the 6th international conference on Language Resources and Evaluation (LREC'08), Marrakech, Morocco", May 2008, http://www.irisa.fr/texmex/publications/versionElect/2008/HGS08a.pdf.

[29] H. JEGOU, L. AMSALEG, C. SCHMID, P. GROS. *Query-adaptative locality sensitive hashing*, in "Proceedings of the 33rd IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP'08), Las Vegas, USA", March-April 2008, p. 825-828, http://www.irisa.fr/texmex/publications/versionElect/2008/jegou08a.pdf.

[30] G. LECORVÉ, G. GRAVIER, P. SÉBILLOT. *An Unsupervised Web-Based Topic Language Model Adaptation Method*, in "Proceedings of the IEEE International Conference on Acoustics, Speech,and Signal Processing (ICASSP'08), Las Vegas, NV, USA", April 2008, p. 5081–5084, http://www.irisa.fr/texmex/publications/versionElect/2008/LeGS2008a.pdf.

[31] G. LECORVÉ, G. GRAVIER, P. SÉBILLOT. *On the Use of Web Resources and Natural Language Processing Techniques to Improve Automatic Speech Recognition Systems*, in "Proceedings of the 6th International Language Resources and Evaluation (LREC'08), Marrakech, Morocco", European Language Resources Association, May 2008, http://www.irisa.fr/texmex/publications/versionElect/2008/LeGS2008b.pdf.

[32] G. MANSON, S.-A. BERRANI. *A TV stream macro-segmentation system*, in "Proceedings of the IEEE International Symposium on Multimedia, Berkeley, California, USA", demo paper, December 2008.

[33] G. MANSON, S.-A. BERRANI. *An inductive logic programming-based approach for TV stream segment classification*, in "Proceedings of the IEEE International Symposium on Multimedia, Berkeley, California, USA", December 2008.

[34] P. NGUYEN-KHANG, A. MORIN, P. GROS. *Boosting for Factorial correspondence analysis for image retrieval*, in "Proceedings of the 6th International Workshop on Content-Based Multimedia Indexing (CBMI'08), London, United-Kingdom", June 2008.

[35] P. NGUYEN-KHANG, A. MORIN, P. GROS. *Caviz, An interactive graphical tool for image mining*, in "Proceedings of the 30th International Conference on Information Technology Interfaces, Cavtat, Croatia", June 2008, p. 371-376.

[36] P. NGUYEN-KHANG, A. MORIN, P. GROS. *Visualization and Interactive Exploration of Factorial Correspondence Analysis Results on Images*, in "Proceedings of the 3rd International Conference on Human Centered Processes , Part I: Main Conference, Delft, The Netherlands", Telecom-Bretagne, June 2008, p. 65-80.

[37] P. NGUYEN-KHANG, A. MORIN, P. GROS, L. QUYET-THANG. *Factorial Correspondence Analysis for image retrieval*, in "Proceedings of IEEE International Conference on Research, Innovation and Vision for the Future (RIVF'08), Ho Chi Minh City, Vietnam", July 2008, p. 269–275.

[38] A. NÉVÉOL, S. E. SHOOSHAN, V. CLAVEAU. *Automatic Inference of Indexing Rules for MEDLINE*, in "Proceedings of the ACL Workshop BioNLP'08, Columbus, OH, USA", short paper, June 2008, p. 88–89, http://www.aclweb.org/anthology-new/W/W08/W08-0612.pdf.

[39] A. OLAFSSON, B.-Þ. JÓNSSON, L. AMSALEG. *Dynamic Behavior of Balanced NV-Trees*, in "Proceedings of the 6th International Workshop on Content-Based Multimedia Indexing (CBMI'08), London, United-Kingdom", June 2008.

[40] P. TIRILLY, V. CLAVEAU, P. GROS. *A Language Modeling Approach to Image Classification*, in "Proceedings of the Networked and Electronic Media Summit (NEM'08), Saint-Malo, France", September 2008.

[41] P. TIRILLY, V. CLAVEAU, P. GROS. *Language modeling for bag-of-visual words image categorization*, in "Proceedings of the international conference on Content-based Image and Video Retrieval (CIVR'08), Niagara Falls, Canada", ACM, July 2008, p. 249-258, http://delivery.acm.org/10.1145/1390000/1386388/p249-tirilly.pdf?key1=1386388&key2=0797619121&coll=ACM&dl=ACM&CFID=82406075&CFTOKEN=46013490.

National Peer-Reviewed Conference/Proceedings

[42] S. HUET, G. GRAVIER, P. SÉBILLOT. *Un modèle multi-sources pour la segmentation en sujets de journaux radiophoniques*, in "Actes de la 15ème conférence sur le Traitement Automatique des Langues Naturelles (TALN'08), Avignon, France", June 2008, p. 49-58, http://www.irisa.fr/texmex/publications/versionElect/2008/HGS08b.pdf.

[43] G. LECORVÉ, G. GRAVIER, P. SÉBILLOT. *Vers une adaptation thématique non supervisée de modèles de langage - utilisation d'Internet comme un corpus ouvert*, in "Actes des 27èmes Journées d'étude sur la parole

(JEP'08), Avignon, France", June 2008, p. 149-152, http://www.irisa.fr/texmex/publications/versionElect/2008/LeGS2008c.pdf.

[44] P. NGUYEN-KHANG, A. MORIN, P. GROS. *Recherche d'images par l'analyse factorielle des correspondances*, in "Actes de la COnférence en Recherche d'Information et Applications (CORIA'08), Trégastel, France", March 2008, p. 23–38.

[45] P. NGUYEN-KHANG, A. MORIN. *Nouvelle approche pour la recherche d'images par le contenu*, in "Actes de la conférence EGC'08, Sophia-Antipolis, France", January-February 2008.

[46] A. NÉVÉOL, V. CLAVEAU. *Apprentissage artificiel de règles d'indexation pour MEDLINE*, in "Actes de la conférence Traitement automatique des langues naturelles (TALN'08), Avignon, France", June 2008, p. 300-309, http://www.irisa.fr/texmex/people/claveau/publis/Neveol-Claveau-TALN08.pdf.

[47] Q.-N. TRAN, T.-N. DO, F. POULET, P. NGUYEN-KHANG. *Vehicle License Plate Classification Using SIFT Descriptors and SVM*, in "Proceedings of the 11th National Conference on Information and Communication Technology (NCICT'08), Hue, Vietnam", 2008.

Scientific Books (or Scientific Book chapters)

[48] V. CLAVEAU. *Translation of Biomedical Terms by Inferring Rewriting Rules*, in "Information Retrieval in Biomedicine: Natural Language Processing for Knowledge Integration", V. PRINCE, M. ROCHE (editors), to appear, chap. 6, IGI - Global, 2009.

[49] M. DELAKIS, G. GRAVIER, P. GROS. *Stachostic Models for Multimodal Video Analysis*, in "Multimodal Processing and Interaction. Audio, Video, Text", P. MARAGOS, A. POTAMIANOS, P. GROS (editors), Multimedia Systems and Applications Series, vol. 33, Springer, 2008, p. 91–109.

[50] S. HUET, G. LECORVÉ, G. GRAVIER, P. SÉBILLOT. *Toward the Integration of Natural Language Processing and Automatic Speech Recognition: Using Morpho-syntax and Pragmatics for Transcription*, in "Multimodal Processing and Interaction. Audio, Video, Text", P. MARAGOS, A. POTAMIANOS, P. GROS (editors), Multimedia Systems and Applications Series, vol. 33, Springer, 2008, p. 201–218.

[51] P. MARAGOS, P. GROS, A. KATSAMANIS, G. PAPANDREOU. *Cross-Modal Integration for Performance Improving in Multimedia: A Review*, in "Multimodal Processing and Interaction. Audio, Video, Text", P. MARAGOS, A. POTAMIANOS, P. GROS (editors), Multimedia Systems and Applications Series, vol. 33, Springer, 2008, p. 3–48.

[52] F. POULET, T.-N. DO. *Interactive Decision Tree Construction for Interval and Taxonomical Data (invited chapter)*, in "Visual Data Mining - Theory, Techniques and Tools for Visual Analytics", S. SIMOFF, M. BOEHLEN, A. MAZEIKAS (editors), Lecture Notes in Computer Science, vol. 4404, Springer-Verlag, 2008, p. 123-135.

[53] F. POULET. *Towards Effective Visual Data Mining with Cooperative Tools (invited chapter)*, in "Visual Data Mining - Theory, Techniques and Tools for Visual Analytics", S. SIMOFF, M. BOEHLEN, A. MAZEIKAS (editors), Lecture Notes in Computer Science, vol. 4404, Springer-Verlag, 2008, p. 389-406.

[54] W.-A. VOGLOZIN, G. RASCHIA, L. UGHETTO, N. MOUADDIB. *From User Requirements to Evaluation Strategies of Flexible Queries in Databases*, in "Handbook of Research on Fuzzy Information Processing

in Databases", J. GALINDO (editor), vol. 1, Hershey, PA, USA: Information Science Reference, 2008, p. 115–142.

Books or Proceedings Editing

[55] L. AMSALEG, V. CLAVEAU (editors). *Actes de la COnférence en Recherche d'Information et Application (CORIA'08) et des Rencontres Jeunes Chercheurs en Recherche d'Information (RJCRI'08)*, Presses de l'Universités de Rennes 1, Trégastel, France, March 2008.

[56] A. KEMPER, P. VALDURIEZ, N. MOUADDIB, J. TEUBNER, M. BOUZEGHOUB, V. MARKL, L. AMSALEG, I. MANOLESCU (editors). *Proceedings of the 11^{th} International Conference on Extending Database Technology (EDBT'08)*, ACM International Conference Proceeding Series, vol. 261, ACM, Nantes, France, March 2008.

[57] P. MARAGOS, A. POTAMIANOS, P. GROS (editors). *Multimodal Processing and Interaction. Audio, Video, Text*, Multimedia Systems and Applications Series, vol. 33, Springer, 2008.

[58] F. POULET, B. LE GRAND, T.-N. DO (editors). *Actes du $6^{ème}$ Atelier Visualisation et Extraction de Connaissances, Extraction et Gestion de Connaissances (EGC'08)*, January 2008.

[59] B. LE GRAND, F. POULET (editors). *Visualisation et extraction de connaissances*, RIA, Revue d'Intelligence Artificielle, Numéro spécial sur la visualisation et l'extraction des connaissances, vol. 22, n^{o} 3-4, Hermès-Lavoisier, 2008.

References in notes

[60] S. WERMTER, E. RILOFF, G. SCHELER (editors). *Connectionist, Statistical and Symbolic Approaches to Learning for Natural Language Processing*, Lecture Notes in Computer Science, Vol. 1040, Springer Verlag, 1996.

[61] G. AMATI, C. J. VAN RIJSBERGEN. *Probabilistic models of information retrieval based on measuring the divergence from randomness*, in "ACM Transactions on Information Systems", vol. 20, n^{o} 4, 2002, p. 357–389.

[62] L. AMSALEG. *Indexation multidimensionnelle*, in "L'indexation multimédia. Description et recherche automatiques", P. GROS (editor), Hermes, 2007, p. 215-244.

[63] J. BAI, J.-Y. NIE. *Using Language Models for Text Classification*, in "Proceedings of the Asia Information Retrieval Symposium, Beijing, China", Oct 2004.

[64] S.-A. BERRANI, L. AMSALEG, P. GROS. *Recherche par similarités dans les bases de données multidimensionnelles : panorama des techniques d'indexation*, in "Ingénierie des Systèmes d'Information", vol. 7, n^{o} 5/6, 2002.

[65] C. BUCKLEY, G. SALTON, J. ALLAN. *Automatic Retrieval With Locality Information Using SMART*, in "Proceedings of the Text Retrieval Conference (TREC)", 1992, p. 59-72.

[66] C. BÖHM, S. BERCHTOLD, D. KEIM. *Searching in High-dimensional Spaces: Index Structures for Improving the Performance of Multimedia Databases*, in "ACM Computing Surveys", vol. 33, n^{o} 3, September 2001.

[67] T. DEAN, K. KANAZAWA. *A model for reasoning about persistence and causation*, in "Articial Intelligence Journal", vol. 93, n° 1, 1989.

[68] R. FAGIN, R. KUMAR, D. SIVAKUMAR. *Efficient similarity search and classification via rank aggregation*, in "Proceedings of the ACM SIGMOD Conference, San Diego, CA", 2003.

[69] L. FLORACK, B. TER HAAR ROMENY, J. KOENDERINK, M. VIERGEVER. *General Intensity Transformation and Differential Invariants*, in "Journal of Mathematical Imaging and Vision", vol. 4, n° 2, 1994, p. 171-187.

[70] A. GIONIS, P. INDYK, R. MOTWANI. *Similarity Search in High Dimensions via Hashing*, in "Proceedings of the 25th International Conference on Very Large Data Bases, Edinburgh, Scotland, United Kingdom", September 1999, p. 518–529.

[71] A. GUTTMAN. *R-Trees: A Dynamic Index Structure for Spatial Searching*, in "ACM SIGMOD", 1984.

[72] C. HARRIS, M. STEPHENS. *A Combined Corner and Edge Detector*, in "Proceedings of the 4th Alvey Vision Conference", 1988, p. 147-151.

[73] T. HOFMANN. *Unsupervised Learning by Probabilistic Latent Semantic Analysis*, in "Journal of Machine Learning", vol. 42, n° 1-2, 2001, p. 177–196, http://portal.acm.org/citation.cfm?id=599631.

[74] H. JEGOU, H. HARZALLAH, C. SCHMID. *A contextual dissimilarity measure for accurate and efficient image search*, in "Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, CVPR'07, Minneapolis, Minnesota, USA", June 2007, p. 1–8.

[75] D. G. LOWE. *Distinctive image features from scale-invariant keypoints*, in "International Journal of Computer Vision", vol. 60, n° 2, 2004, p. 91–110.

[76] K. MURPHY. *Dynamic Bayesian Networks: Representation, Inference and Learning*, Ph. D. Thesis, University of California, Berkeley, 2002.

[77] X. NATUREL. *Structuration automatique de flux vidéos de télévision*, Ph. D. Thesis, University of Rennes 1, April 2007, http://www.irisa.fr/centredoc/publis/theses/irisafolder.2007-03-01.1633154603/irisapublication.2007-04-26.0496201623.

[78] D. NISTÉR, H. STEWÉNIUS. *Scalable recognition with a vocabulary tree*, in "Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, CVPR'06, New-York, New-York, USA", vol. 2, June 2006, p. 2161–2168.

[79] M. OSTENDORF. *From HMMs to Segment Models*, in "Automatic Speech and Speaker Recognition - Advanced Topics", chap. 8, Kluwer Academic Publishers, 1996.

[80] L. PAULEVÉ. *Réseaux geométriques pour la recherche et l'indexation en grandes dimensions*, Technical report, n° 1903, IRISA, september 2008.

[81] L. RABINER, B.-H. JUANG. *Fundamentals of speech recognition*, Prentice Hall, Englewood Cliffs, NJ, 1993.

[82] S. ROBERTSON, S. WALKER, S. JONES, M. HANCOCK-BEAULIEU, M. GATFORD. *Okapi at TREC-3*, in "3rd Text Retrieval Conference", 1995, p. 109–126.

[83] J. ROBINSON. *The K-D-B-Tree: A Search Structure For Large Multidimensional Dynamic Indexes*, in "ACM SIGMOD", 1981.

[84] G. SALTON, C. BUCKLEY. *Term-weighting approaches in automatic text retrieval*, in "Information Processing and Management", vol. 24, n^{o} 5, 1988, p. 513–523.

[85] G. SALTON. *Automatic Text Processing*, Addison-Wesley, 1989.

[86] C. SCHMID, R. MOHR. *Local Grayvalue Invariants for Image Retrieval*, in "ieee Transactions on Pattern Analysis and Machine Intelligence", vol. 19, n^{o} 5, May 1997, p. 530–534.

[87] J. SIVIC, B. C. RUSSELL, A. A. EFROS, A. ZISSERMAN, W. T. FREEMAN. *Discovering Object Categories in Image Collections*, in "Proceedings of the International Conference on Computer Vision, Beijing, China", 2005.

[88] J. SIVIC, A. ZISSERMAN. *Video Google: A Text Retrieval Approach to Object Matching in Videos*, in "Proceedings of the International Conference on Computer Vision", vol. 2, October 2003, p. 1470–1477.

[89] M. UTIYAMA, H. ISAHARA. *A Statistical Model for Domain-Independent Text Segmentation*, in "Proceedings of the 39th Annual Meeting of Association for Computational Linguistics, ACL'01, Toulouse, France", July 2001.

[90] C. J. VAN RIJSBERGEN. *Information Retrieval, 2nd edition*, Dept. of Computer Science, University of Glasgow, 1979.