Bringing Interactive Contextual Maps to Users for Information Retrieval

Viaud Marie-Luce[1], Renoust Benjamin[1, 2], Agnès Saulnier[1], Jérôme Thièvre[1]

1 : INA, Bry sur Marne, France
2 : CNRS UMR5800 LaBRI & INRIA Bordeaux - Sud Ouest, Talence, France

E-mail: {mlviaud, brenoust, asaulnier, jthievre}@ina.fr

Abstract: **The growth of available online collections of documents comes with a growth in users' expectations for retrieval services. Contextual representations and interaction processes are becoming a real challenge to accurately navigate among large amounts of data. This paper presents an intuitive framework for the creation of multilevel interactive maps of multimedia content. Our framework allows the user to build interactive 2D representations of multimedia documents. The visualization is based on a graph layout and filtering and generates maps through document clustering. Users may interactively modify the map according to their criteria of interests, defining their own viewpoints on the collection. They can then dig or surf depending on the interaction mode they choose. Those functionalities allow them to navigate, browse, and analyze quickly and precisely sets of multimedia documents, allowing deep investigation of collections.**

Keywords: HMI, information visualisation, interaction, multimedia retrieval, similarity search, relevance feedback, graph clustering

1 INTRODUCTION

The French Audiovisual Institute (Ina), in charge of storing and preserving the French audiovisual heritage, needs to ensure its exploitation and to make it more readily available. Under the terms of the French law voted on 20 June 1992, the Inathèque de France is responsible for collecting and preserving radio and television broadcasts. To achieve this mission, 150 media librarians annotate the audiovisual documents collected daily, at different levels of precision. Ina's archive holds 1 500 000 pictures and 1 300 000 hours of video (600 000 digitized) and more than one million documentary notes. From 1998 to 2010, Inathèque de France has increased the scope of its collections and 100 TV channels and 20 radio channels are being collected. The entire collection of Ina's archives is made available online for professionals by subscription via www.inamediapro.com. Since April 2006, Ina's public web site has assembled more than 20000 hours of Radio and TV broadcasts, representing about 85000 programmes or excerpts.

Documenting these ever-growing contents with constant human resources makes the use of new technologies a necessity. A study on Ina's professional clients shows that digitization of video contents leads to new uses and needs. Clients expect more precise and diverse access to content, with a finer description of the resources. Needs for new tools are then clearly expressed to assist archivists in documenting resources and to provide users with new types of search and navigation in video contents. In Ina's professional context, semi-automatic tools are promoted to assist archivists in their tasks without loss of control, allowing them to focus on human high added value, which is to produce coherent and normalized annotations at high semantic level.

On the other hand, the Ina online site addresses the general public, which opens new paradigms for resource access. Discovering TV archives becomes a recreation process: search strategies are totally different, mainly based on the same rules as web search. Users will follow advice from friends, have butterfly navigation behaviors in the resources or try loose queries. So, open interfaces for navigation and discovery are required.

This paper presents few cognitive elements that explicit the mental processes and strategies involved in information retrieval. Section 2 focuses on the principle of uncertainty of the search process to develop the importance of context and polyrepresentation of the world from both the user side and data contents. We present the answer of HMI and Visualization for retrieval tasks and discuss it according to the previously introduced notions. Related works on user interaction for Information Retrieval (IR) are presented in Section 3. Section 4 describes our framework for analyzing and visualizing multimedia collections. Finally, Section 5 presents our interface prototype, introducing a few applications with users evaluation. The conclusion discusses some open issues and presents our future work.

2 A FEW COGNITIVE ELEMENTS FOR IR REPRESENTATION AND INTERACTION

Many cognitive studies have aimed to analyse the cognitive processes involved in information retrieval [1, 2, 3]. Ellis develops six characteristic behaviors relative to information retrieval: starting, chaining, browsing, differentiating, monitoring, and extracting. Marchionini underlines the multiple goals of search activities, from a basic lookup task to a complex domain investigation. He observes that the growth of available data, thanks to

Corresponding author: Benjamin Renoust, INA 4 av de l'Europe 94366 Bry sur Marne Cedex FRANCE, brenoust@ina.fr

online search on the web or specific collections, makes user expectation of services beyond lookup much higher. A deeper analysis of complex cognitive search process is developed by Ingwersen, who outlines the concept of polyrepresentation applied to both the user cognitive space and the information space of IR systems. Moreover, the multiplicity of the possible behaviors associated to the polyrepresentation of resources and to the act of interpretation makes the process of information retrieval uncertain and unpredictable. In order to give better clues for perception and interpretation and reduce the uncertainties and the unpredictability, efforts to provide users with representations and interaction modes are then needed on both sides of the communication channel. The high dynamicity of cognitive structures together with the faculty of analysing overlap between user viewpoints play a fundamental role in our world perception. The possibility to express this dynamicity through interaction and representations becomes therefore a challenge for information seeking.

HMI and information visualisation address this issues [4, 5, 6]. Our work in these fields tries to emphasize the polyrepresentations of data collections through both global and local approaches.

3 RELATED WORKS

As shown by Lew et al. [7], we can find a wide panel of different approaches that allow the user to retrieve multimedia documents depending on the context of her search and the context provided by the data. Beyond the traditional (even advanced) text query, we focus our interest on map generations that help the user to locate her query among amounts of results reflecting the data's context. We focus also on user interactions, which provide way to control queries and to manipulate the data.

NAV-GRAPHE [8] is a hierarchical graph-based navigation process of multimedia documents. The software describes and indexes documents linking them by similarities. Given a starting node and its neighbors, the user can choose an image (representative of the document) that will be used as a starting node for another navigation step if needed. This hierarchical interface presents a partial view of the navigation graph to the user: only a sample of images is presented at each step of the retrieval process.

Google Image Swirl [9] organizes images in a structured tree from a partition of a search results into hierarchical clusters created from a multimodal description of the documents. The interface presents a radial layout in which each node selected has its children spread around. The user can navigate through this tree of up to a thousand images keeping the mental path she followed being able to go back to any step previously chosen. A similar approach, but applied to video content, named MediaGraph is offered by Ina to the public in its website [10].

Worring et al's MediaMill video search engine [11] proposes four browsing tools. Among them The Galaxy Browser interests us the most. It uses 2D similarity-based visualization of image frames where the user can annotate a large collection of images. The system improves the efficiency of its maps through a relevance feedback process allowing the user to manipulate the images in the displayed space in order too improve their classification.

In diverse ways, the systems described above allow the user to narrow her search in a wide pool of documents by generating maps taking into account the user's perspective in a context defined by the set of documents displayed.

The Tulip software [12] offers a huge framework for graph visualization and analysis. It allows the user to input a graph and process it with different layouts, selections, and measures. The graph can be also partitioned and even hierarchies can be constructed among the subgraphs created. This allows the user to display, explore, and analyse simultaneously different views of the same graph in order to reach or highlight specificities of the graph. The user can then create her own perspective of the data given multiple views.

Microsoft Research Asia improved the traditional query suggestion through a scheme named Visual Query Suggestion [13]. The suggestion complements the user keyword with extra keywords as query suggestion usually does but adding to each a collection of images that gives visual clues. The images are also used in the corresponding query to refine the results. As *a picture is worth a thousand words*, this approach escorts the user in her search helping her to precise her query with visual information.

4 MULTIMODAL MAPS GENERATION: FRAMEWORK DESCRIPTION

4.1 Framework overview

In the following, we are calling *multimedia document* or *entity* a media resource (still image, video or radio segment) associated with its textual description. Textual descriptions may present different levels of precision. The shortest review contains only a few factual fields like "title", "channel"," date" or "time", while the most descriptive ones, such as TV news programs, are described with precise textual content: keywords, people and location names and free text summaries.

Our framework contains different types of modules:

- Several *indexation modules* analyze each resource to extract cross modal low-level features. These processes are executed off-line. Specific files containing the low-level global or local descriptors are generated and associated to the original resource files. Mono-modal or cross-modal distance matrices are computed from the descriptors.
- An *index structure and search* module allows time efficient searches in huge sets of indexes (1 billion 20D-descriptors). In fact, features extraction generates a huge amount of indexes. There is a trade-off between the potential quality of the retrieval and

the quantity of indexes which describe the resources.

- A Graphic User Interface module, which provides users with interactive visualizations to get an overview, to analyze or navigate in sets of structured or non-structured items. The size for the document set is for the moment limited to 35000 visualized entities.

We have chosen a web services technology which allows both of these capabilities: several teams may work on different components of the global system at the same time; the efforts for the integration steps must be minimized. The integration of several distributed components in a common application consists in our case in building a web user interface which consumes various services and creates a specific representation of the results.

4.2 Media description and similarity measures

We have developed several modules of extraction of global visual low level features such as histograms in different colour spaces (RGB, LUV and HSV), gradient orientation histogram, corner detection or motion estimation [14]. Visual local descriptions are based on Harris or SIFT features and Symmetry points [15]. The framework also provides accurate and flexible tuning during the search process: the rate of extracted local features can be changed on-line, as well as the granularity of the targeted video segments can be adapted to the application.
However, we are also using efficient results from other teams for audio, text and visual indexes. In the context of The European project VITALAS, visual indexes and similarities are built by the IMEDIA/INRIA [16] team. More complex transmodal similarity measures are computed from these descriptions [17, 18].

4.3 Similarity map generation

Our visual representation is based on graph models [19]. The modelling of graph structures has several advantages:

• First, the paradigm used for graph layout matches our objective in terms of representation: the proximity between entities and groups of entities in the 2D representation should best reflect their proximities in the N-dimensional descriptor space.

• The mathematical duality between graphs and matrices makes graphs efficient for expressing distance matrices.

• Graph filtering and layout have been already largely studied and drawing techniques exist, more or less adapted to our data [20].

• Passage from valued to symbolic data may be easily performed on the model. Links removal may promote topology emergence in the data set.

Our approach uses a customized energy force model algorithm [21, 22] to achieve the layout. Within this model, we consider a repulsive force between nodes, and a spring force between connected nodes. Each edge is seen as a spring characterized by its resting length and stiffness coefficient. For our applications, the goal of the visualisation is to identify neighborhoods and aggregates. Multi Dimensional Scaling is not very efficient in layout quality and computing time because we have to consider the complete graph. Then, we have implemented generic filtering methods of nodes and edges based on their inner or topological attributes (centrality, degree, hub/authority values…). For graphs based on similarity matrices, the highest similarities of each node appears to be the most significant to elaborate the layout. Then, we generate a graph issued from KNN-mutual filtered matrices. In such cases, the radius of a cluster and the distance between two clusters are related to the inverse of the edge density (normalized edge-cut) [23]. Filtering functionalities may be adjusted interactively to make the view more readable.
Finally, we use a standard agglomerative hierarchical clustering algorithm to identify and label clusters in the rendering space. To obtain clusters of arbitrary shapes, we choose a linkage metric based on the minimum distance between objects. These distances can be parameterized in the interface, allowing the user to *control* the view.
Current TF/IDF techniques are used to extract the most significant word for each cluster, allowing the labelling of clusters.

The layout, done iteratively in real time, works on sets of 100,000 documents and about 1,000,000 edges.

5 USER FEEDBACK AND DATA MANIPULATION

5.1 General presentation of the interface

The interface presents 3 panels: the left side presents an overview of the area of the map currently viewed, and a detailed view of the document pointed at the mouse; the middle window shows the map; the functional panel is on the right (see **Figure 1**). The map view shows documents and clusters based on the previously described similarity. Several mouse actions are available for the user:

- Zoom: use of the mouse wheel to zoom in and zoom out.

- Translation: right button to translate the map

- Selection: left button to select items or groups of items

- Positioning the mouse on a cluster while pushing the *control* key displays its most salient terms in the labelling panel.

The functional panel contains:

- Search and labelling functionalities and parameters: Text input area may be written directly or set by selecting terms of the current cluster label set. Results are highlighted with colorful squares in the map and displayed on a list in the panel (**Figure 2**).

- Layout and visibility functionalities and parameters: results items of a search are highlighted. Sets of items may be selected on the map with the mouse and marked by a change of colors (Figure 2, 3, and 4).

- Map creation: A new map may be created from a set of selected or highlighted items by selecting the "map creation" function in the main menu (Figure 5).

5.2 The Wikipedia "tourism" case

The *Wikipedia "tourism" map*[*] (Figure 1 and Figure 2) has been built with Xerox textual and visual descriptions. A textual document description gathers page title, paragraph title and caption of the document where it exists. Visual descriptions are based on global descriptions of the image. Cross similarities are computed to exploit simultaneously image and text proximities [18]. The map gathers more than 50000 documents. Two ways to investigate the map's content are offered to users: moving the mouse on clusters to display their textual content (this mode allows to *surf* on the map). Zooming allows the observation of an area or a specific cluster, documents are displayed on the map according to the level of zoom. Textual or visual queries may be performed, and results are marked on the map.

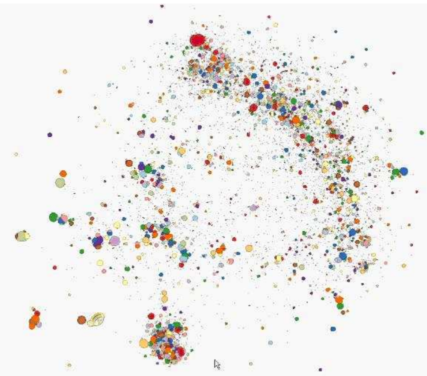


Figure 1: The global map of *Wikipedia "tourism"* images. Images and textual similarities have been combined. The map presents two main parts: the bottom left part gathers synthetic images (flags, icons, drawings, schemes…) while the upper right crescent gathers real images.

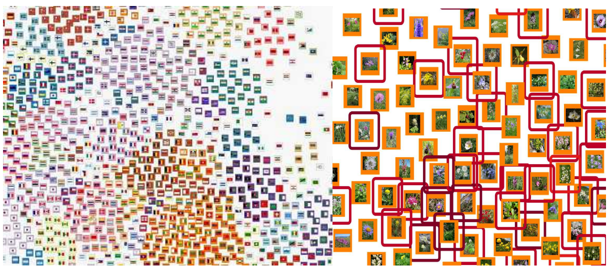
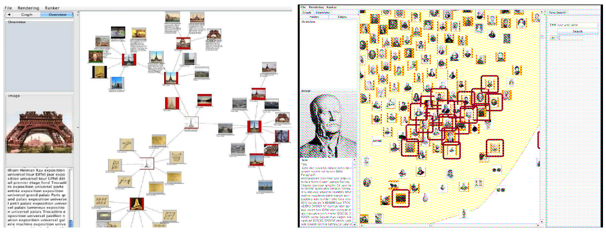


Figure 2: Example of clusters for the global map: the flag cluster and the flower cluster. Flower cluster has been found from the query "flower". The results containing the word flower are marked showing that all the documents of the cluster do not contain the word.

5.3 Combining local and global exploration with advanced relevance feedback

In [24], the authors propose a system that couples serendipitous browsing and query-based search in a smooth manner. The proposed system offers two levels, one global and one local, of visualizing the context of the information seeking task. It also allows to view and search the data using either monomodal or cross-modal similarities. Furthermore, the system integrates a new relevance feedback model that takes into account the multimodal nature of the data in a flexible way and a combination of two parameters being the locality and a forgetting factor, that allows the user to design adaptive metrics in the interactive search process (Figure 3).



(a) (b)

Figure 3: Screenshots of the system for the Eiffel Tower scenario of *Wikipedia "tourism"*: a) a local map in the middle (with different developed search directions) and b) an example of a subtopic viewed from the global map on the right.

5.4 Reorganizing data from the user's viewpoint

Intermediate representations or submaps may be built from results or data selection directly on the map. New clusters are generated, creating a new viewpoint on a sub-collection. This process illustrates the potential of the polyrepresentation of the resource descriptions and allows enhancing of the users knowledge perception, according to the available data content. Selection may concern any of the metadata description level, such as date, personality, themes, type of programs or even modality. The example of Figure 4 and Figure 5 shows the dedicated context for the query "Mitterrand" in the Ina collection. Documents about "Mitterrand" are selected and a specific sub map is computed. The submap creation is automatically created in real time. Few parameters have still to be set to achieve the layout. Then, clusters are computed, labelled, and rendered. The submap allows users to see quickly while discovering the cluster content what type of resources concerning Mitterrand may be found in this collection and outlines his main activities according to TV diffusion. The main clusters are about socialist political party, elections, foreign policy, presidential visits, inauguration, and political actions on specific conflicts.

[*] Thanks to the French INFOMAGIC project

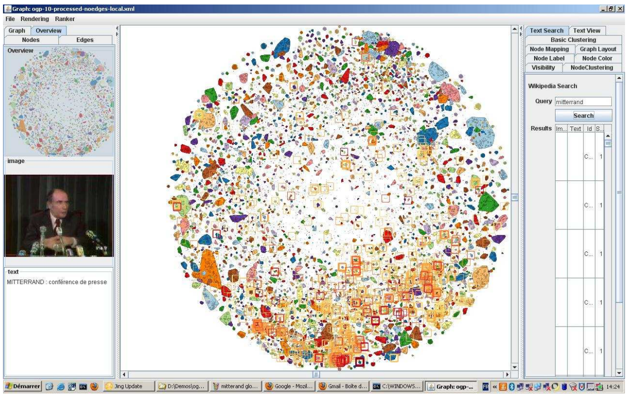


Figure 4: Screen shot of the global map (~60000 video documents available on Ina web site) organised in semantic clusters based on description and keyword similarity. The query "Mitterand" is marked by colored squares mapping the rank of the retrieval from red to yellow. Results are distributed mainly on the bottom of the map which gathers political themes.

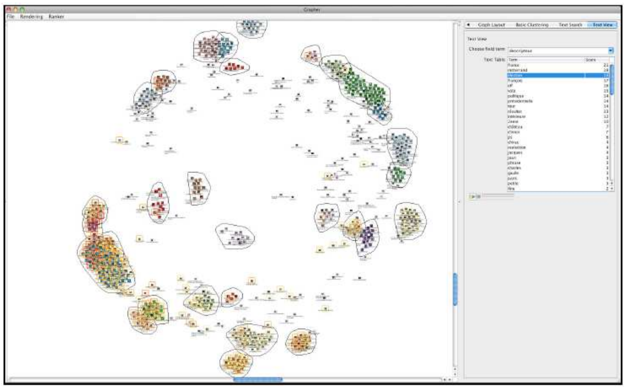


Figure 5: Submaps: Reorganisation of the Mitterrand results: clusters reflect the main TV subjects about Mitterrand: elections, foreign policy, presidential visits, inauguration, and specific political actions.

5.5 User evaluation

User evaluations have been conducted on these three prototypes with five Ina professional archivists [†]. The evaluation [25] relies on utility and usability criteria:

- Utility: theoretical evaluation

- Usability: pluralistic walkthrough to study effectiveness and user satisfaction. Pluralistic walkthrough focuses on qualitative results.

The goal of the evaluation was to test the use of our interface to browse, navigate and analyse a "substantial" set of contents. More specifically, we wanted to test if these representations and interactions allow users to acquire a better knowledge on the data and THUS develop search strategies with better control. The pluralistic walkthrough is an extension of the cognitive walkthrough that includes real users. Each evaluation was done in two hours. Each user was asked to develop three scenarii of information seeking, containing search and analysis of results according to their completeness and diversity.

This method is based on discussions and comments while achieving a scenario of use. Creators of the tool conduct the manipulation. The user comments the achievements of the tasks from his point of view. The discussion should answer the following questions:

Q1. Does the user know what to do to achieve the task?

Q2. Is it easy for users to understand how to achieve the tasks with the interface?

Q3. Does the user understand the results given by the system?

Q4: Does the visualisation and manipulation enhance its knowledge on the data available?

The compilation of the evaluation results shows that:

Q1: Collection exploration by moving the mouse on the clusters and textual search entry are intuitive and no major problems have been encountered. Zoom and translation functions are effective to explore a detailed area.

Q2: No major problem has been encountered to achieve the tasks

Q3: Exploration focuses intuitively on the most important clusters and groups of clusters of the map. Some users asked why some queries appear very sparse on the map.

Q4: "The map is nice to show an overview of the content. No existing tool gives such view". The display of keywords shows the descriptive vocabulary and facilitates query reformulation and data understanding.

Submaps make easiest the analysis of results, especially when the set of results is too big or sparse to be parsed efficiently on the global view. However, creation of submaps should be made totally automatic, the process is still too difficult to handle by novice users.

On the combination of local and global exploration more details in [24] allows the user to better control the exploitation/exploration trade-off:

- The local map allows users to express iteratively and more precisely their information needs while the global map allows to better understand the different boundaries of their search and to discover non-expected subtopics.

- The use of "forgetting" and locality factors are encouraged though we should not lose the user by asking her to tune a lot of parameters.

Recommendations focuses on giving the number of documents in a cluster and the creation of meta clusters to make the discovery even more efficient.

6 CONCLUSION

We have presented several solutions to interactively search and explore collections of contents allowing users to get multiple representations of the data. Users are given multiple ways of interaction with the data, either by

† Thanks to INFOMAGIC and VITALAS European project.

modifying their viewpoints, or by using different modes of navigation (dig or surf) to reach theirs goals. Multilevel representations (global, intermediate and local views) add to context perception, helping users situate themselves. As analysed in [3] human knowledge is based on the overlap of the polyrepresentations of the “world”, and the possibility offered to the user to explore data from multiple viewpoints constitutes the first experiment that instantiates this process. However continuity and coherence of the maps represent a real issue in data manipulation and representation. On global maps, hierarchical representation of clusters, allowing gathering clusters on higher level themes is another issue put forward by our users.

The maps presented in this work are all static and don’t include any notion of time or viewpoint transition. As we are especially interested by archives, most content can be described over time or in a dynamic mater that depends on a continuous parameter. Indeed, the historical context of a document is also a key element while digging into archives. Exploring dynamic data, according to multiple viewpoints or time consideration, comes naturally as our next step. Moreover, to answer Ina’s users recommendations, we are also seeking for hierarchies of document clusters, eventually to pull out hierarchies of time clusters.

References

[1] D. Ellis, (1993) "A Behavioural Approach to Information Retrieval System design" Journal of Documentation, Vol. 45 Iss: 3, pp.171 - 212

[2] Marchionini, G.: Exploring search: from finding to understanding. Communications of the ACM 49 (2006) 41–46

[3] P. Ingwersen, "Cognitive Perspectives of Iinformation Retrieval Interaction: Elements of a cognitive theory", Journal of Documentation, Vol. 52 Iss: 1, pp.3 - 50, 1993

[4] Ware, C. Information Visualization: Perception for Design, ISBN 1558608192, 2004

[5] Card, SK. Mackinlay, JD. Shneiderman, B. Readings in information visualization: using vision to think, ISBN 1-55860-533-9, 1999

[6] Keim, DA. Information visualization and visual data mining, 2002

[7] Lew, MS. et al, Content-based multimedia information retrieval: State of the art and challenges, ACM Transactions on Multimedia Computing, Communications, and Applications (TOMCCAP), 2006

[8] Don, A. Indexation et navigation dans les contenus visuels: approches basées sur les graphes, Ph. D. Thesis, Université de Bordeaux, France, 2006

[9] Jing, Y. et al. Google Image Swirl: A Large-Scale Content-Based Image Browsing System, 2009.

[10] http://www.ina.fr (launched in 2009)

[11] Nguyen, GP. Worring, M. Interactive access to large image collections using similarity-based visualization, Journal of Visual Languages & Computing, 2008

[12] Auber, D. Tulip-a huge graph visualization framework, Graph Drawing Software, 2003

[13] Zha, Z.J. et al. Visual Query Suggestion, Proceedings of the seventeen ACM international conference on Multimedia, 2009

[14] H Goeau, O Buisson and ML Viaud, "Image collection structuring based on evidential active learner", at CBMI, London, 2008

[15] A. Joly, O. Buisson and C. Frélicot "Content-Based Copy Retrieval Using Distorsion-based Probabilistic Similarity Search",published in IEEE Transactions on Multimedia, 2007.

[16] N. Boujemaa and J. Fauqueur and M. Ferecatu and F. Fleuret and V. Gouet and B. Le Saux and H. Sahbi", IKONA for interactive specific and generic image retrieval", MMCBIR 2001, France", 2001.

[17] N. Grira, M. Crucianu, and N. Boujemaa. Fuzzy, clustering with pairwise constraints for knowledge-driven image categorization. In IEEE, Proceedings on Vision, Image and Signal Processing,2006

[18] S. Clinchant, JM Renders and G Csurka, " Advances in Multilingual and Multimodal Information" Lecture Notes in Computer Sciences, Book, p569-576, Vol5152, Springer 2008

[19] Thièvre J. PHD Thesis, INA/Univ of Montpellier, 2006

[20] I. Herman, G. Melançon, and M. S. Marshall. Graph visualization and navigation in information visualization: a survey. IEEE Transactions on Visualization and Computer Graphics, 6(1):24-43, 2000.

[21] A. Noack. An energy model for visual graph clustering. In G. Liotta, editor, 11th International Symposium on Graph Drawing (GD 2003), LNCS 2912, pages 425{436, Berlin, 2004. Springer-Verlag.

[22] T. M. J. Fruchterman and E. M. Rheingold, “Graph drawing by force directed placement,” in Software - Practice and Experience, 1991, vol. 21, pp. 1129–1164.

[23] A. Noack. Energy-based clustering of graphs with non uniform degrees. In 14th International Symposium on Graph Drawing (GD 2006), pages 309-320, Limerick, Ireland, 2006.

[24] J. Ah Pine, JM Renders, ML Viaud, "A Continuum between Browsing and Query-based Search for User-Centered Multimedia Information Access" proceedings of AMR 09, to be published by Springer

[25] deliverable of european VITALAS project