

CoViz: cooperative visualization to facilitate sense making by groups of users

Bérenger Arnaud, Guillaume Artignan, Jérôme Cance, Gabriel Delmas,
Mountaz Hascoët, Nancy Rodriguez

► **To cite this version:**

Bérenger Arnaud, Guillaume Artignan, Jérôme Cance, Gabriel Delmas, Mountaz Hascoët, et al..
CoViz: cooperative visualization to facilitate sense making by groups of users. IV'11: 15th International Conference on Information Visualisation, Jul 2011, London, United Kingdom. pp.255-260, 2011, <10.1109/IV.2011.48>. <lirmm-00808117>

HAL Id: lirmm-00808117

<https://hal-lirmm.ccsd.cnrs.fr/lirmm-00808117>

Submitted on 4 Apr 2013

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

CoViz: cooperative visualization to facilitate sense making by groups of users

Bérenger Arnaud, Guillaume Artignan, Jérôme Cance, Gabriel Delmas, Mountaz Hascoët, Nancy Rodriguez
LIRMM, UMR 5506 CNRS
Univ. Montpellier II
Montpellier, France
{arnaud, artignan, mountaz, nancy.rodriguez}@lirmm.fr

Abstract

Coviz is a tool that uses visualization to support spontaneous construction, organization and exploration of collections. We use the term collection to identify a set of documents gathered for a given purpose by one or several individuals. Coviz facilitates creation, organization and update of collections by providing multi-scale visualization and interaction techniques. Coviz has been tested in several informative case studies involving small groups of users. In these situations, Coviz has fulfilled several important aspects of both groups and personal organization needs. Our initial experiments indicate that Coviz can even encourage unusual and interesting forms of cooperation among people.

Keywords--- Visualization, collaborative sense making, spontaneous coordination.

1. Introduction

Over the past few years numerous research efforts have been made to help people to search and organize the information they have to handle every day. Web search, web navigation, sense making, annotating, organizing, recommending and sharing have all been studied intensively. Advanced solutions have been proposed for each of these information management related activities. However, combining complementary yet separated approaches in a transparent and seamless way seems necessary to support knowledge emergence and is still a challenging problem.

Our objective with Coviz is to provide a tool supporting a broad range of information management activities. Based on advanced interaction and visualization techniques, Coviz integrates searching, bookmarking, annotating, harvesting, recommending, clustering, classification and sharing.

Coviz is designed for relatively small groups of users involved in cooperative sense making from web resources. Even though recent efforts have brought to the scene tools that support for collaborative web search, these approaches do not solve the problem of sense making and even sometimes they increase its complexity as noted by Paul and Morris [13]. Indeed, supporting

collaborative sense making implies supporting very challenging issues such as group decision making, multiple search coordination, and merging of multiple personal information spaces, organization styles, and practices.

A key issue in organizing information spaces is to find a way of conciliating both group and personal perspectives that may differ widely. An interesting study by Abrams and Baecker [4] investigated the ideal organization of personal information spaces. Meanwhile, a huge literature has investigated group information organization. As noted by Halpin et al. [10], “optimal means by which to organize information often pits formalized classification against distributed collaborative tagging systems”. Group needs and especially consistency and stability over time is well embodied in formalized classification whereas distributed collaborative tagging systems are better suited to personal organization needs such as reactivity, spontaneity and rapidity. Our approach in Coviz consists in trying to conciliate both group and personal requirements for the organization of collections of documents and collaborative sense making.

The major contribution of this paper consists in (1) a transparent interaction model that combines different visual analytics techniques to facilitate creation, organization and update of collections and (2) informative experiments of the use of Coviz in real-world situations.

2. Coviz overview

Coviz handles textual data and metadata coming from heterogeneous sources (search engine, harvesters, personal web references, etc) and shared among a group of users involved in information management related tasks. Data handled in Coviz is mainly made of documents, uniquely identified by URLs. Coviz metadata is considered as a set of dimensions along which Coviz data can be annotated. Coviz is limited to a static set of dimensions to better focus on the design of the appropriate user interface for cooperatively handling several different dimensions. This has to be considered as an initial step necessary to reduce the complexity of the problem, but our longer term objective is to handle any arbitrary set of dimensions to represent any relevant metadata.

From the observation of user practices we have determined which initial dimensions could be chosen as mostly relevant. Amongst all possible dimensions, we chose dimensions that suit the following requirements: genericity, discrimination power, orthogonality, heterogeneity, simplicity and utility. The resulting set of preliminary chosen dimensions is (1) tags, (2) people, (3) collection, (4) quality ratings, (5) free text comments, (6) document kind. These dimensions are generic enough and were found useful for most data collections handled in Coviz experiments. The dimensions are very different in nature which makes them complementary to adequately annotate the data.

Tags can be seen as flat lists, or can be organized as hierarchies. Since all users handle their own tags personally, Coviz is designed to handle a set of hierarchies along which data can be annotated. This

People dimension is simpler since it is represented by a flat list of nominal values consisting of the name of commentators or authors referenced in the system.

Quality rating is a simple single-valued numerical dimension.

Free-text comment is a single-valued unstructured text dimension similar to description or people comments.

Document kind is a single valued categorical dimension similar to the content-type of mime metadata but while mime content-type describes low-level categories of documents, Coviz types of documents are higher-level categories such as research paper, talk, course, thesis, etc. As with other dimensions, the purpose of this dimension is to meet the requirements above for supporting the design of the user interface and in term to be replaceable by any other categorical dimension.

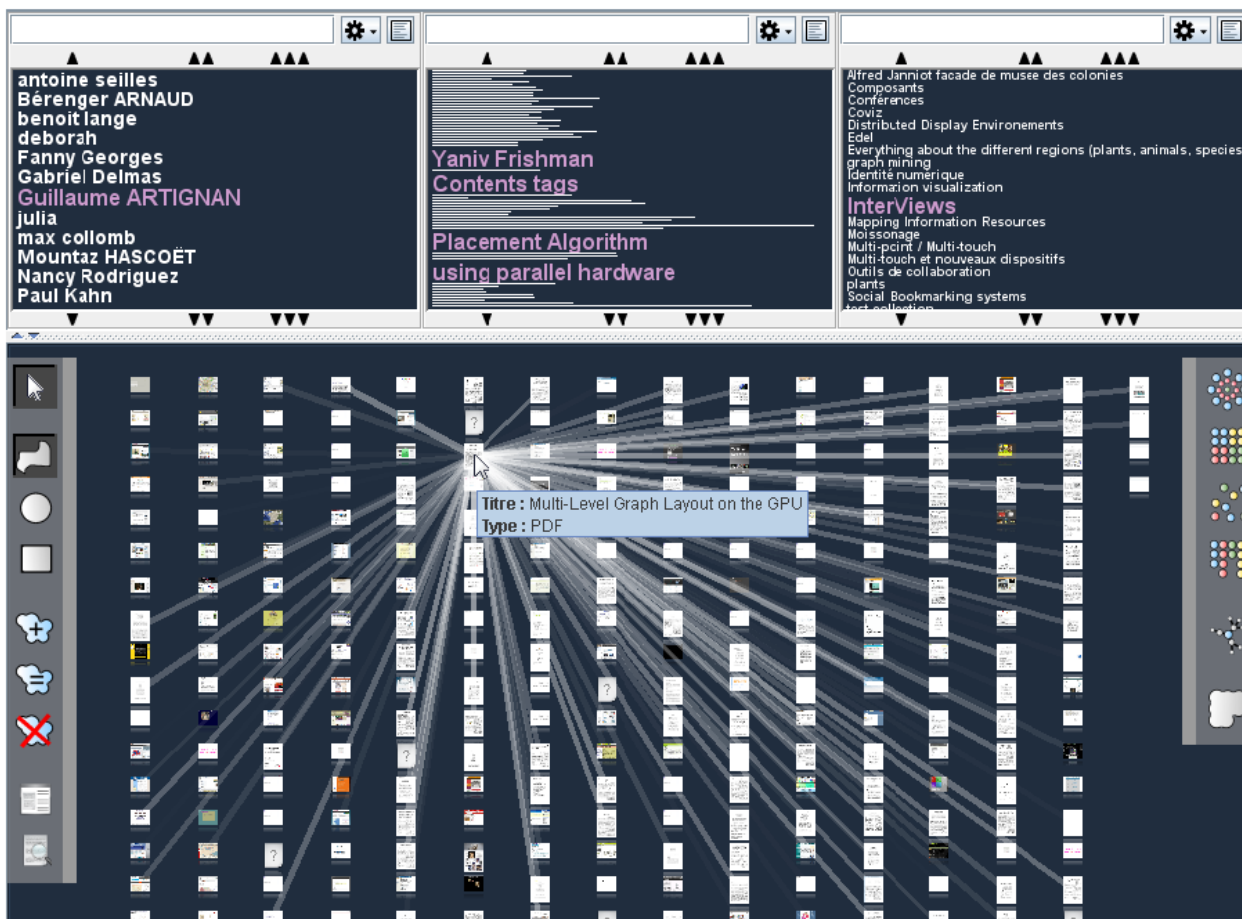


Figure 1: Coviz visualization and interaction window

dimension is also interesting because it is a multi-valued dimension. Several tags can be associated explicitly with documents or can be implicitly deduced. As a result the tag dimension is both multi-hierarchical and multi-valued.

Collection name can be considered as a multi-valued dimension that aims at providing direct access to the documents contained in collections. This dimension can be considered as an articulation or transversal dimension used to make it possible to support reuse of browsing results and combination of faceted browsing with other types of information gathering. The specificity

of this dimension can be formally presented. First, let's recall how a set of data is defined. By extension it consists in enumerating the items in the data set. By intension it consists in exhibiting the set of features which are shared by the data contained in that set and that are sufficient to identify it. The specificity of the dimension collection is that it helps to define subsets of data both in extension, and in intension, whereas all other dimensions can only retrieve sets in intension. The experiments showed that this dimension was very useful for a user eager to merge subsets of "manually" selected documents with otherwise intentionally defined collections.

3. Graphical coding in Coviz

The visual representation of document is based on the graphical encoding of a multivariate clustered graph [3]. Nodes represent the Coviz documents, and edges represent a similarity relation between documents. Each node contains the following attributes: an identifier, a title, a URL, a date of insertion and a date of modification.

Edges are computed from the similarity between nodes – e.g. documents. Documents are represented in a vector space using Term Frequency (TF) and Inverse Document Frequency (IDF) [13] to compute their coordinates in that space. TF corresponds to the frequency of apparition of a word in the document. IDF corresponds to the frequency of apparition of a word in the entire corpus of documents. The TF-IDF measure corresponds to the ratio of these two measures. A Pearson's correlation is further computed between vectors representing documents in order to compute their similarity [14].

Graphically, documents are represented by thumbnails. Links are represented by lines between documents. The more the color is opaque and the more similar are the documents. Links between documents are visible on mouse over. A bubble shape surrounding the documents represents the clusters. Mouse allows direct actions on documents while keys shortcuts help supporting interaction for sets.

The graphical encoding is further implemented using STOOG [3], a graphical graph drawing API. The main originality of STOOG is the explicit declaration of style sheets for graph visualization. Fig. 1 illustrates the resultant visualization of documents in Coviz.

4. Tightly coupled fisheye lists for faceted metadata browsing

One important issue with faceted browsing [11] is the curse of dimensionality [7]. Coviz addresses this problem in three complementary ways: (1) a visualization of dimension values combined with advanced browsing and multi-level scrolling, (2) multi-scale visualization of documents, collections and relations and (3) tight coupling between (1) and (2).

Another important issue with faceted browsing is the exploration of the interrelations between dimensions. This type of exploration is quasi impossible for a user given the default user interfaces, like those found on most faceted based web site, for example. Rarely cited in the literature this apparently minor problem has, in fact, significant and negative impact on both exploration of data and overall sense making from it. Indeed, missing cues about how values are distributed and related across dimensions in a given data set constitutes a real hindrance in understanding what the content of a collection is. We call this problem the lack of *metadata insight*.

We have addressed the lack of metadata insight by combining two different modes of faceted browsing in Coviz. The first mode is similar to the default mode where the selection of items in the fisheye list views of dimension values filters the data in the document view (cf. Figure 1) the item selected are displayed in purple. The selection of items on different dimensions is interpreted as a conjunction of disjunction when the dimensions are multi-valued and a simple conjunction when they are single-valued. We call this mode the selective mode that directly filters the data.

The second mode of browsing is called the highlight mode and was designed to address the lack of metadata insight problem. It consists in highlighting related values in all dimensions when the user moves over items in the fisheye list. For example, if the mouse goes over a specific collection in the fisheye list displaying collections, all the tags found in the data of that collection will be highlighted in the fisheye list displaying tags whatever other selected criterions. As for tags, all other values appearing in the data will be highlighted in the other fisheye lists displaying other dimensions. This highlighting can be seen as a very simple type of brushing and was found so useful by users that one user asked to save these highlights. Enriching interaction in order to capture dynamic highlighting and transform it into a selection is in the roadmap of our future work.

5. Multi-scale visualization and clustering

Two kinds of clustering are proposed in Coviz, the interactive clustering and the automatic clustering.

The interactive clustering consists in letting the user construct its own clusters of documents. The documents must be selected from one or several groups. The selection is the result of a search or a manual selection. Fig. 3 presents a sample of selection. The documents colored in blue, red and green are respectively in three selection groups. The execution of the clustering could transform these three groups of selection in three clusters. A cluster is considered as an element which can be moved or removed.

The automatic clustering takes the graph of documents and returns the clustered graph. We use the MCL algorithm [5] for clustering our weighted graph.

The algorithm is based on Markov chains and random walks [5].

6. Architecture and implementation

Coviz system is implemented with Web technologies. The system is composed by (1) a server which providing services data storage, generation of dynamic pages and extra computation, and (2) a client part supporting visualization and interaction. In this section we present the system architectures and our implementation choices.

Architecture

The server has three components: database, web server and application server. The database enables data storage such as online documents, user account and tag hierarchies. It can be queried using the SQL relational language. The Web server provides generation of HTML pages and connection to database by the means of PHP language. The application server provides thumbnail generation and computes all necessary transformations on the data. This second server is powered over Java. Results are cached in database for future queries.

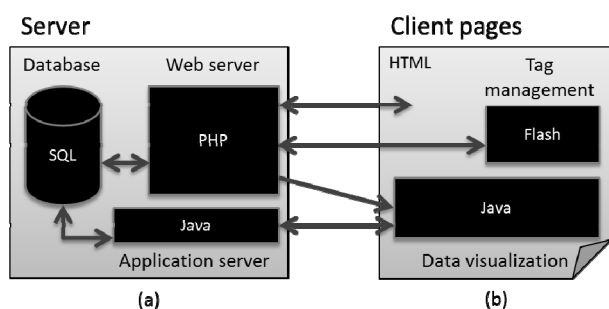


Figure 2: Overview of Coviz architecture

The client part consists mainly of HTML pages which is enough to do basic interaction and visualization. To avoid the basic limitation, web pages embedded two rich internet applications: (1) Flash modules, written with Adobe Flex language, for tag handling, and (2) a Java Applet for the multi-scale visualization of documents.

The figure 2.a presents the server architecture and 2.b the client architecture.

Implementation choices

Given the architecture properties, we made choices for implementing the user-oriented folksonomy approach and for keeping the user-centered context. Our choices impact features such as tags and documents meta-data.

An important issue concerns implicit tagging mechanism implementation, extracted from explicit tagging and sharing. We consider two options: (1) physically storing implicit tags or (2) computing implicit tag on-the-fly. The first option requires checking implicit tag integrity each time a tag is updated. The second

option is also costly because it requires processing each time a tag is selected. Finally, we choose the first option that favors fast response for visualization and data scaling.

7. Experiments and lessons learned

We have experimented Coviz in two different contexts: (1) a research project involving approximately 20 persons including researchers, engineers, designers, and PhD students, (2) a entire class of 30 undergraduate students. We describe the context, the objectives, the problems and lessons learned in this section.

In the first experiment, Coviz was used as an experimental tool and our objective was mainly to test very different user interfaces to best suit user needs. Several prototypes of the user interface based mainly on hierarchical list displays were initially designed to support sense making related tasks in a collection of documents coming from the social science community. These preliminary user interfaces were favored by people that were exposed to them, over alternative designs found in more popular systems. However, we found these preliminary user interfaces very frustrating because it did not meet two objectives: (1) a fast and fluid way to interact both at the item level and at the collection level and (2) a good way to conciliate visualization and interaction for groups and visualization and interaction for individuals. Clearly, even with advanced drag-and-drop based interaction, hierarchically displayed lists were far from meeting the challenges of handling both items and collections in a fluid interaction model. Secondly, another frustration came from the current state of the art of most frequently used default roles found in social systems that determine who is entitled to create/edit/delete what resource. Models that are based mainly on permission rights appeared as a real breaking point for the emergence of any significant spontaneous, fluid and cooperative interaction model useful for collaborative sense making related tasks.

Concerning the challenges of the visualization of both items and large collections, we designed the last version of Coviz (cf. Figure 1) by combining of (1) graph visualization, (2) cluster visualization, (3) faceted browsing and (4) fisheye lists displays. This last version was used successfully and was way closer to meet the requirements we had than previous ones. However, it was obvious from usage that it still suffers from minor usability pitfalls mostly due to technological limitations and easy to fix in the future versions. Another lesson was learned from this experiment with the last version. The experiment raised interesting questions about the direct manipulation model of faceted based browsing. The combination of fisheye list display, multi-grain scrolling, highlighting and selecting modes of faceted browsing and the coupled graph based visualization was found to support fluid interaction pretty well according to our criteria. However, several filtering interaction strategies were tested and more experiments would be useful to

adjust those strategies to better suits personal and group needs.

This version of Coviz was experimented to support resource gathering and annotating in a class. For this second experiment, students were undergrads and Coviz has provided help to them with their last and most important assignment. For this assignment students had to work in a both cooperative and competitive mode. The competitive mode was that they were asked to work by groups of two and each group had to perform the same project. The project consisted in creating a 3Dinteractive and graphic scene involving visual representations of Nao, a humanoid from Aldebaran [9] and reproducing in a virtual scene all the moves that Nao could perform for real. The cooperative mode consisted in using Coviz to gather and annotate all the resources necessary to achieve the projects. We made it clear from the beginning that both the administration and usage of Coviz were open to all students. Any student could add/remove/edit any resource. The only two simple guidelines were announced before they started: (1) if the system breaks you are all responsible and will have to fix it and (2) only high quality resources should be included and individuals are responsible for the quality of the resource they include in the system.

By using Coviz, and these very simple indications, the 30 students made an impressive job at collectively gathering high quality resources very rapidly. There was an impressive diversity in the ways that lead different students to different high quality resources. In less than one hour, a set of very highly relevant resources were collected and made available for all students at once. Our experiment with using other approaches in other classes is that this collaborative resource gathering takes several weeks without ever reaching that level of quality. This result was very encouraging and even beyond our initial expectations. Based on students comments and observations our speculation is that the next important step is to investigate how the editing and scaling factors should be adjusted. Indeed, one important problem was that no student dared comment the resources of the others. They spoke about them but did not write about them. We speculate that this is because the annotation tool was too formal and definitive to suit the need for politeness or solidarity among students. One student suggested that a forum be associated with Coviz so that casual discussion about the resources can be separated from the set of annotated resources.

8. Related work

Knowledge emergence from social networks

The emergence of knowledge from social networks either directly or after analysis is a very active field of research and has gained popularity with the apparition of social web sites such as Delicious, CiteUlike or others. In this area, systems, algorithms and studies have been proposed in a variety of domains including collaborative web search[7] , collaborative sense making [13]

collaborative tagging systems, recommendation systems [1] , etc. These research efforts have mainly focused on providing either the analysis of used systems and design recommendations for the next generation or systems that perform collaborative information and knowledge related tasks. Our focus in Coviz is very different from those approaches as our aim is to provide a transparent, consistent and flexible user interface that can handle most information or knowledge management tasks in a consistent and spontaneous way. As a consequence, Coviz features in terms of search features, bookmarking services or data analysis are minimal. At the same time, our experience with the inconsistency between systems that perform best in each area is maximal and the focus of Coviz is to help with conciliating the different approaches.

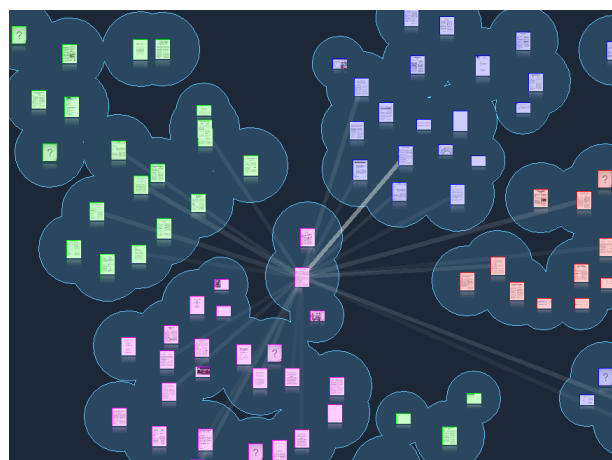


Figure 3: Clustering of data using Coviz

Farooq et al. [6] have analyzed two years of CiteUlike data and proposed 6 metrics for a better understanding of tagging behaviors. Their results indicate that (1) the number of tags used in CiteUlike is consistently growing as well as the number of new users. This model of tag growth is exactly the opposite of the results found by Marlow and David reporting a study with Flickr data. With Flickr data, tag growth is diminishing as time goes by. One speculation to conciliate these contradictory results is that tag growth is a consequence of growth of new users. This explanation would be consistent with the well-known vocabulary problem [8] , which states that the probability that two different users use the same label to describe the same concept is very low. The study of Farooq et al. brings to the scene very interesting questions and ends with recommendations for the next generation of social tagging systems. One of them is to facilitate tag reuse and our own study of user behavior corroborates this finding, so our approach in Coviz accounted for this recommendation carefully. Farooq et al. also insisted on recommending tags that are informationally powerful – e.g. with a high discrimination power and a high level of non-obviousness [6] . We found this recommendation more debatable considering our observations of users at

work and their tagging motivations. We speculate that such a strategy would only result in creating a scale free network e.g. the distribution of tag usage would be a stabilized power law similar to what can be found in [10]. Our experiment in designing Coviz showed the importance of conciliating two conflicting usages of tags: the personal and the group perspectives. Conciliation of personal and group perspectives is achieved by supporting tag sharing over multiple personal tag hierarchies.

Social bookmarking and tagging systems

Social bookmarking principle can be described by a tripartite graph composed of users, bookmarks and web resources and links between those elements.

Since tagging in social bookmarking systems is central and has been extensively studied. Tagging is also included in every document oriented application, in which it is usually considered as one among others metadata. In [12], three kinds of tagging are suggested (1) blind tagging where users do not know others user's tags, (2) visible tagging where the user can see the others user's tags and (3) suggested tagging where the system proposes tags to the user.

However, current social bookmarking systems have some important limitations. The vocabulary problem is one of them and was mentioned previously in this paper. Disambiguating tags when *polysemic tags* are used is another well-known challenging problem. Collaborative construction of consistent tag hierarchies to improve Folksonomies or tag-based ontologies is also very preliminary and often considered as an open issue [16].

While most of the proposed approaches focus on system support of tag processing, our approach in Coviz is to integrate these processes in a visual system and to focus on interaction and cooperation among users. By first letting users construct their individual hierarchies of tags, shared hierarchies are further created progressively from individual ones. A direct benefit of our approach is to avoid the vocabulary problem since users can keep their own vocabulary, and share it explicitly with others while preserving the control over sharing processes. Similarly, multi-lingual tagging, as well as the polysemy problem can benefit from this visual analytics approach and the combination of interaction, visualization and automatic processing.

9. Conclusion

In this paper we have presented Coviz, a system to collaboratively visualize and interact with collections. We have reported lessons learned from real case studies that have shown that the combination of visualization and advanced interaction techniques was a difficult yet fruitful approach to collaborative sense making challenges. The experiments have also shown that meeting these challenges requires a deeper analysis and re-design of models of human cooperation that determines how people share things. Most frequent permission rights based models are clearly not suitable

for fruitful cooperation. The variety of alternative models that have been proposed so far is probably a good indicator of both the need for a change and the size of the challenge to find the next model.

Acknowledgements

Many thanks to Benoit, Antoine, Paul and all the people that provide fruitful comments and help in achieving this project. This work is partially funded by the TGE Adonis, from the French national research funding programs. Avec le Soutien du TGE Adonis.

References

- [1] Adomavicius, G.; Tuzhilin, A. "Toward the Next Generation of Recommender Systems: A Survey of the State-of-the-Art and Possible Extensions", IEEE Transactions on Knowledge and Data Engineering 17 (6): 734-749, 2005.
- [2] Amershi, S. and Morris, M.R. CoSearch: A System for Co-located Collaborative Web Search. Proceedings of ACM CHI, ACM Press 2008.
- [3] Artignan, G., Hascoët, M.: STOOG - Style-Sheets-based Toolkit for Graph Visualization. ICEIS'10, Portugal. 2010.
- [4] David Abrams, Ron Baecker, and Mark Chignell. 1998. Information archiving with bookmarks: personal Web space construction and organization. In Proceedings of the SIGCHI conference on Human factors in computing systems (CHI '98).
- [5] Dongen, S. V. Graph Clustering by Flow Simulation. PhD thesis, University of Utrecht. 2000.
- [6] Farooq, U., Kannampallil, T. G., Song, Y., Ganoe, C. H., Carroll, J. M., and Giles, L. 2007. Evaluating tagging behavior in social bookmarking systems: metrics and design heuristics. In Proceedings of the 2007 international ACM GROUP '07, 351-360, 2007.
- [7] Freyne, J. and Smyth, B. Cooperating search communities. In Proc. Adaptive Hypermedia and Adaptive Web-based Systems 2006, 1--10.
- [8] Furnas, G. W., Landauer, T. K., Gomez, L. M., and Dumais, S. T. The vocabulary problem in human-system communication. Commun. ACM 30, 11 1987.
- [9] Gouaillier David, Hugel Vincent, Blazevic Pierre, Kilner Chris, Monceaux Jérôme, Lafourcade Pascal, Marnier Brice, Serre Julien, Maisonnier Bruno: Mechatronic design of NAO humanoid. ICRA 2009
- [10] Halpin, H., Robu, V., and Shepherd, H. The complex dynamics of collaborative tagging. WWW '07. ACM, New York, NY, 211-220. 2007.
- [11] Ka-Ping Yee, Kirsten Swearingen, Kevin Li, and Marti Hearst. Faceted metadata for image search and browsing. CHI '03. ACM, 2003
- [12] Marlow, C., Naaman, M., Boyd, D., and Davis, M. HT06, tagging paper, taxonomy, Flickr, academic article, to read. In Proceedings of the Seventeenth Conference on Hypertext and Hypermedia (Odense, Denmark, August 22 - 25, 2006). HYPERTEXT '06. ACM, 2006
- [13] Paul, S.A., and Morris, M.R. (2009). CoSense: Enhancing Sensemaking for Collaborative Web Search. In Proceedings of the Conference on Human Factors in Computing Systems (CHI 2009), Boston, MA.
- [14] Pearson, K. Mathematical contributions to the theory of evolution. Regression, heredity and panmixia. Philos. Trans. Roy. Soc. London Ser. 1896.
- [15] Sparck Jones, K. A statistical interpretation of term specificity and its application in retrieval. In Document Retrieval Systems. P. Willett, Ed. Taylor Graham Series In Foundations Of Information Science.1988
- [16] Wu, X., Zhang, L., and Yu, Y. 2006. Exploring social annotations for the semantic web. WWW '06. ACM, 2006