

An Adaptive Knowledge Domain Visualization for Conferences

Peter Kraker

Marshall Plan Scholarship Paper

University of Pittsburgh
Graz University of Technology

Prof. Dr. Peter Brusilovsky
Personalized Adaptive Web Systems Lab (PAWS)
University of Pittsburgh

Prof. Dr. Stefanie Lindstaedt
Knowledge Technologies Institute
Graz University of Technology

February-April 2014

Contents

Acknowledgements	ii
Abstract	iii
1 Introduction	1
1.1 Motivation	1
1.2 Previous Work	2
1.2.1 Head Start	2
1.2.2 Conference Navigator	3
1.3 Objectives and Research Questions	4
1.4 Methodology	5
1.5 Overview of the Following Sections	6
2 Altmetrics-based Visualizations	7
2.1 Introduction	7
2.2 Data and Method	7
2.3 Results	9
2.4 Discussion and Future Work	10
3 Adaptive Overview Visualizations	13
3.1 Introduction	13
3.2 Data and Method	13
3.3 Results	17
3.4 Evaluation Plan	19
4 Conclusions and Future Work	21
List of Figures	23
Bibliography	24

Acknowledgments

I would like to thank Prof. Brusilovsky and his team at the Personalized Adaptive Web Systems Lab of University of Pittsburgh for their inspiration, their help and their valuable feedback. They have made me feel very welcome during my stay. I also want to thank Prof. Druzdel and his team for accommodating me in their office.

Sincere thanks go to International Relations and Mobility Programmes Office at TU Graz and the Office of International Services at University of Pittsburgh who assisted me during all stages of the process.

I would like to thank the Marshallplan Foundation for extending a scholarship to me. Additional financial support for my research stay was provided by a Panton Fellowship of the Open Knowledge Foundation. These programs are crucial for early career researchers as they provide much needed international experience. Finally, I would also like to thank my advisor Stefanie Lindstaedt for her continued support and mentoring.

Abstract

Information overload is not a contemporary issue in science. The enormous increase in scientific knowledge has also had an impact on conferences. From small events, many conferences have transformed into huge endeavors with multiple sub-events, simultaneous tracks, and thousands of participants. Therefore, it becomes important to provide conference participants with an overview of conference talks. Beyond that, it is also important to adapt this overview to the interests of conference participants.

In this paper, I present work on an adaptive knowledge domain visualization for conferences. It is based on the knowledge domain visualization Head Start and the conference scheduling system Conference Navigator 3 (CN3). In a first study, bookmarking data from CN3 is used to show the topical structure and the evolution of this structure for the UMAP Conference.

The first results from this type of analysis are encouraging. Using small multiples allows for a comparison of knowledge domain visualizations over various years. This approach, however, also has a weakness limiting its usefulness: being based on user interactions, it can only be created when enough bookmarking data is available. It therefore cannot be used before the conference begins - this, however, is the time when the system is most useful to conference participants.

To overcome this weakness, I performed a second study employing content-based measures instead of bookmarking data. In this study, a co-term approach incorporating external knowledge was developed and applied to the ITS 2014 Conference. In addition, the interface is adapted to the individual user profile. This results in a visualization that lets users see their bookmarked talks and highlights recommended talks for them. In order to validate this visualization, an evaluation is proposed that covers both personalization features and usability aspects.

1 Introduction

1.1 Motivation

Information overload is not a contemporary issue in science. Price (1961, 1963) showed that modern science has been growing exponentially since its inception more than 400 years ago. Figure 1 depicts the most famous graph on development of the number of journals over time. Exponential growth can still be observed for the number of papers (Larsen and von Ins, 2010) and the number of researchers (National Science Board, 2010). This phenomenal growth did not only bring about an unparalleled increase in knowledge, it also led to an omnipresent state of information overload.

To deal with this overload, science has developed a number of instruments that facilitate categorization and evaluation of scientific output:

- Conferences and journals to provide for the collection and exchange of knowledge on a certain topic
- Peer review to weed out non-scientific and unoriginal research
- Quantitative indicators for evaluating scientific output.

The enormous increase in scientific knowledge has also had an impact on conferences (Farzan and Brusilovsky, 2007). From small events, many conferences have transformed into huge endeavors with multiple sub-events, simultaneous tracks, and thousands of participants. The World Wide Web Conference 2014¹ for example featured 6 simultaneous tracks. Of these 6 tracks, the research track alone had 84 talks in 28 sessions; 4 sessions were always held at the same time. In addition, conference participants could choose between 20 workshops, 15 tutorials, 123 posters, 30 demos, 4 panels and 3 keynotes over the course of five days.

As one can see, it is quite hard to get an overview of a conference such as WWW 2014. One needs to work through extensive schedules to determine the sub-events one is interested in. Thematic sessions provide some guidance, but with the sheer amount of sessions and the multi-thematic nature of many papers, it is often not possible to locate all interesting talks. Therefore, it becomes important to provide conference participants with an overview of conference talks. Beyond that, it also important to adapt this overview to the interests of conference participants.

¹<http://www2014.kr>

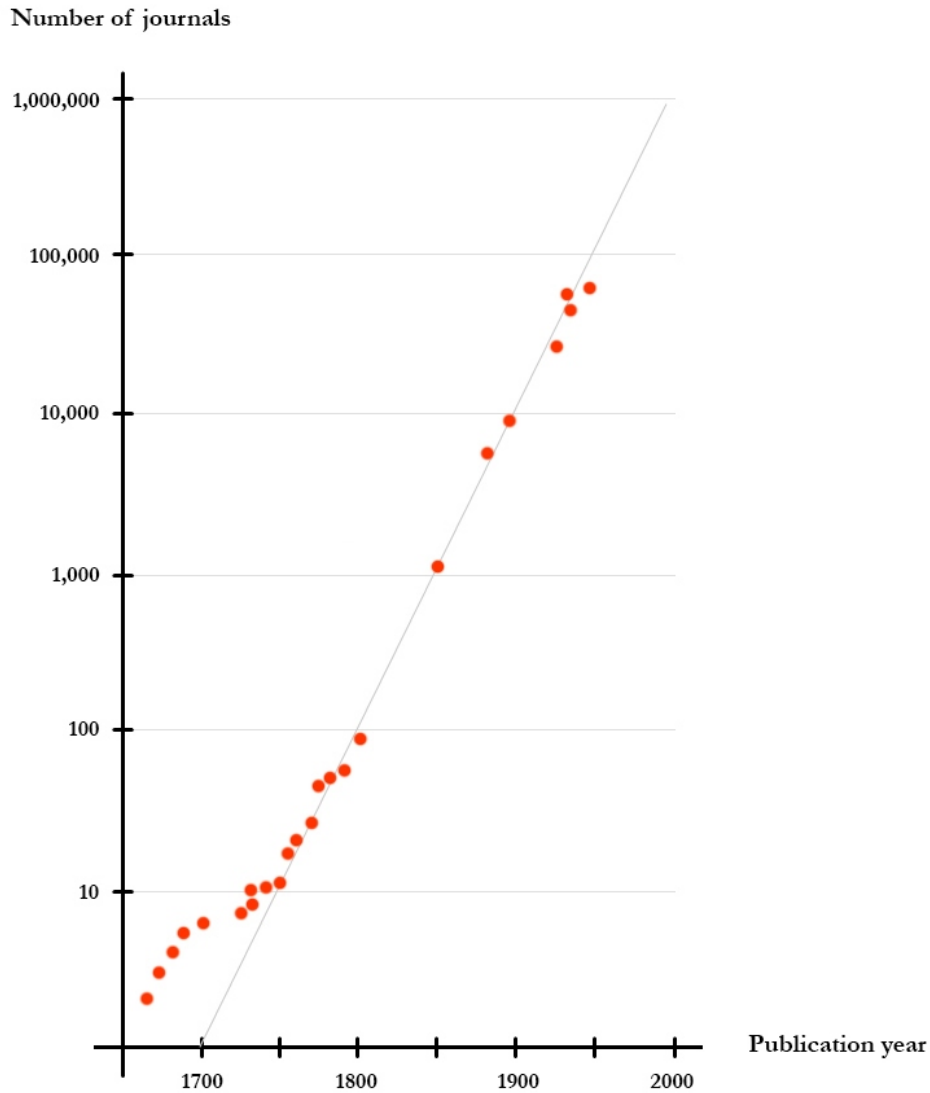


Figure 1: Growth of the number of scientific journals on a logarithmic scale starting from the 17th century. Adapted from Price (1961).

1.2 Previous Work

1.2.1 Head Start

In my previous work, I have shown that readership data can be used to structure a scientific field. In particular, I have studied co-readership as a measure of subject similarity in the online reference management system Mendeley (Kraker et al., 2012). I have then employed readership statistics to analyze the most popular papers in the

field of educational technology (Kraker et al., 2013). This resulted in an interactive visualization which gives an overview of the most important areas in a field, and shows relevant publications related to these areas (Kraker, 2013), see Figure 2.

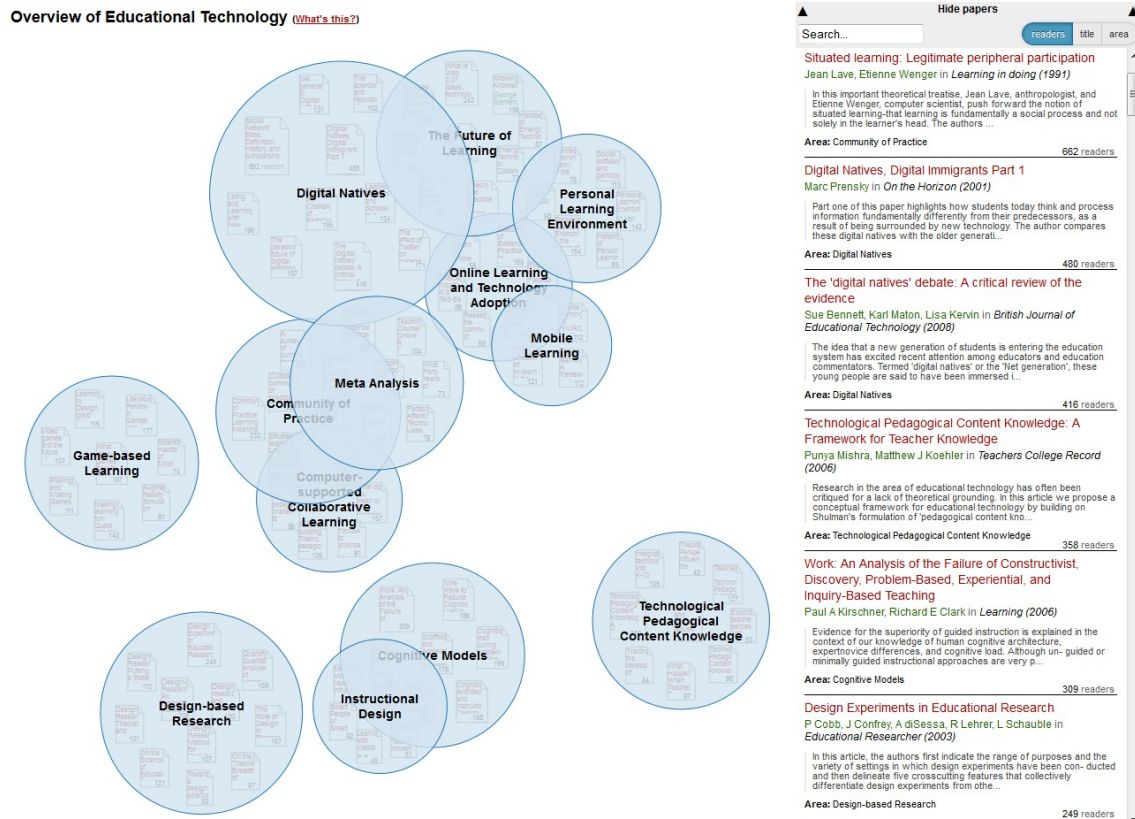


Figure 2: Overview visualization of educational technology based on co-readership

An evaluation showed that the visualization based on readership statistics does provide a diverse and recent representation of the field. Researchers like the idea behind the visualization and would recommend it as a starting point for students in educational technology. Nevertheless, the evaluation also revealed that researchers would like to have a personalized overview. Thus, adaptivity has emerged as a crucial requirement for the system to be effectively used as a learning tool.

1.2.2 Conference Navigator

The Personalized Adaptive Web Systems Lab (PAWS) of the School of Information Sciences at University of Pittsburgh is the leading research group in the world for web-based adaptive learning systems. The group has developed various adaptive sys-

tems based on usage data and overlay user models, such as CoMeT for research talks (Brusilovsky et al., 2010), Progressor for learning resources (Hsiao and Brusilovsky, 2012), and Conference Navigator for conference scheduling (Wongchokprasitti et al., 2010).

The system relevant for this study is Conference Navigator 3 (CN3) (Parra et al., 2012). CN3 allows conference attendees to create a personal conference schedule by bookmarking talks from the program that they intend to follow (see Figure 3). The system also provides users with recommendations based on the talks they have already bookmarked. Furthermore, users can explore the list of attendees and their social media profiles.

The screenshot shows the Conference Navigator 3 interface for UMAP 2013. The top navigation bar includes 'Home', 'Papers', 'People', 'My UMAP', and 'TalkFusion'. The main content area displays a grid of sessions for June 10, 11, 12, 13, and 14. Each session entry includes the time, session name, location, title, authors, and type.

Jun 10	Jun 11	Jun 12	Jun 13	Jun 14
<p>09:00-10:30 Tutorials 1 - Session 1 Location >> Room N10 Design and Evaluation of Recommender Systems ? Bridging the Gap between Algorithms and User Experience Authors: Paolo Cremonesi, Franca Garzotto, Pearl Pu Type: Tutorial</p>	<p>09:00-10:30 EMPIRE 2013 - Session 1 Location >> Room N13 Welcome Time: 2013-06-10 09:00-09:30</p> <p>Building Systems to Capture, Measure, and Use Emotions and Personality Time: 2013-06-10 09:30-10:30 Authors: Neal Lathia Type: Workshop Invited Talk Track: EMPIRE 2013</p>	<p>09:00-10:30 GroupRS 2013 - Session 1 Location >> Room N14 Welcome Time: 2013-06-10 09:00-09:30</p> <p>Sequential Music Recommendations for Groups by Balancing User Satisfaction Time: 2013-06-10 09:30-10:00 Authors: Auste Piliponyte, Francesco Ricci, Julian Koschwitz Type: Workshop Paper Track: GroupRS 2013</p> <p>Group Recommendation in an Hybrid Broadcast Broadband Television Context Time: 2013-06-10 10:00-10:30 Authors: Oskar van Deventer, Joost de Wit, Jeroen Vanattenhoven, Mark Guelbahar Type: Workshop Paper Track: GroupRS 2013</p>		
<p>10:30-11:00 Coffee Break Break</p>				
<p>11:00-12:30 Tutorials 1 - Session 2 Location >> Room N10 Design and Evaluation of Recommender Systems ?</p>	<p>11:00-12:30 EMPIRE 2013 - Session 2 Location >> Room N13</p>	<p>11:00-12:30 GroupRS 2013 - Session 2 Location >> Room N14</p>		

Figure 3: Interface of Conference Navigator 3 showing the schedule of UMAP 2014

1.3 Objectives and Research Questions

The goal of this work is to develop an adaptive knowledge domain visualization of the artefacts contained in Conference Navigator. To achieve this goal, I followed

three objectives:

1. Generation of knowledge domain visualizations based on data from Conference Navigator 3
2. Implementation of personalization features for the visualization
3. Evaluation of the personalized overview visualization

From these three objectives, I derived the following research questions:

- RQ 1:** What are adequate indicators derived from Conference Navigator for creating knowledge domain visualizations?
- RQ 2:** What are relevant personalization features for overview visualizations of conferences?
- RQ 3:** How to evaluate the adaptive knowledge domain visualization of a scientific conference?

1.4 Methodology

To achieve the objectives outlined in the above section, I followed a four step procedure:

1. Analysis of the properties of readership and other indicators in Conference Navigator: I analyzed relevant system properties in order to determine a set of indicators that are suitable for visualization.
2. Visualization of topical structures as a browser-based interface: In this step, I adapted the current Head Start visualization to Conference Navigator.
3. Implementation of adaptive features: In this step, I extended the visualization to adapt to the user profile by honoring past activities and preferences of the user.
4. Planning of an evaluation of the visualization with users of Conference Navigator: In this step, I devised an evaluation plan for the visualization. The goal of the evaluation is to find out, how usable the visualization is and to what extent it helps users to find interesting talks more efficiently.

1.5 Overview of the Following Sections

The following sections are structured as follows: in section 2, I describe work on a knowledge domain visualization spanning multiple years of a conference based on bookmarking data in Conference Navigator. In section 3, an adaptive knowledge domain visualization based on article metadata and externally sourced keywords is presented. In addition, an evaluation plan is proposed for this visualization. Finally, in section 4, I draw conclusions and present avenues for future research.

2 Almetrics-based Visualizations Depicting the Evolution of a Knowledge Domain

2.1 Introduction

Almetrics have recently received a lot of attention in the scientometric community and beyond. A lot of effort is going into assessing the potential of altmetrics for evaluative purposes. Yet another strand of research has emerged, focusing on altmetrics for relational purposes. One problem that is considered within the latter strand of altmetrics research is knowledge domain visualization. So far, clicks (Bollen and van de Sompel, 2006) and readership (Kraker et al., 2013) have been successfully employed to map a scientific domain. These efforts, however, represent the state of a domain, but not the evolution of a domain.

The evolution of scientific domains has been addressed in the past primarily using citation-based analysis; see e.g. Garfield et al. (1964) and Small (1999). There is, however, a significant problem with citations: it takes around two to six years after an article is published before the citation count peaks (Amin and Mabe, 2000). Therefore, citation-based visualizations, and indeed all analyses that are based on incoming citations, have to deal with this time lag. Altmetrics have thus emerged as a potential alternative to citation data. Compared to citation data, altmetrics have the advantage of being available earlier, many of them shortly after the paper has been published. In many instances, usage statistics are also easier to obtain and collect (Haustein and Siebenlist, 2011).

In this section, work towards showing the evolution of scientific domain using data from a scientific conference scheduling software is presented. Conference Navigator Parra et al. (2012) allows conference attendees to create a personal conference schedule by bookmarking talks from the program that they intend to follow. This scheduling data represents an altmetrics source which – to the best knowledge of the author – has not been studied before.

2.2 Data and Method

All data is sourced from Conference Navigator 3. As a use case, I have chosen to analyse the 19th and 20th iteration of the Conference on User Modelling, Adaptation and Personalization (UMAP)², representing the conference years of 2011 and 2012. The procedure for creating knowledge domain visualizations follows the approach used in the knowledge domain visualization Head Start which employs Mendeley

Parts of this section have been submitted as Kraker P., Weißensteiner P., Brusilovsky, P. (2014): Almetrics-based Visualizations Depicting the Evolution of a Knowledge Domain. *19th International Conference on Science and Technology Indicators*.

²<http://www.um.org/conferences>

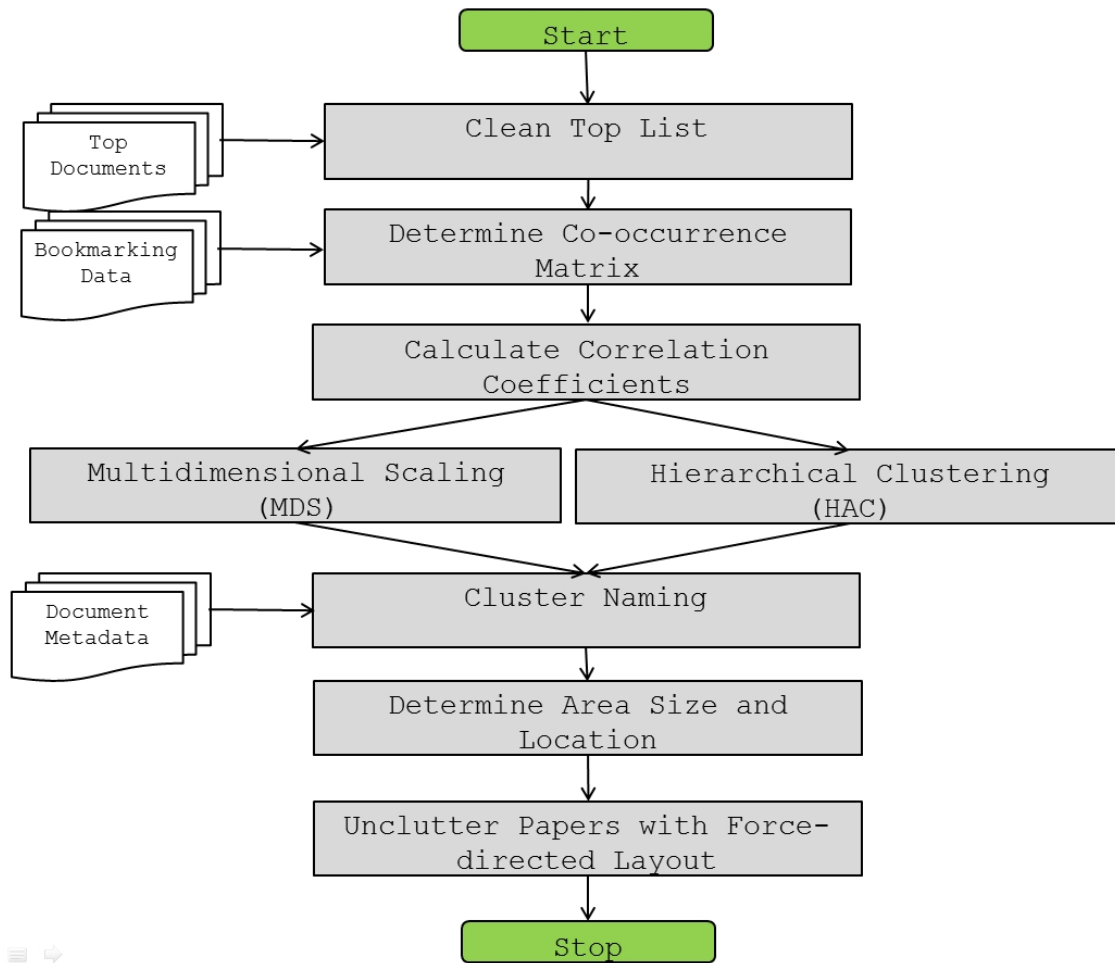


Figure 4: Knowledge domain visualization process

readership data to map a research field Kraker et al. (2013). It is adapted from the knowledge domain visualization process described in Börner et al. (2003).

This process is shown in Figure 4. At first, individual knowledge domain visualizations were created for each year. Therefore, the publication metadata and the scheduling data have been extracted from the system ranked by the number of bookmarks received. A threshold of 4 (3) bookmarks was introduced for 2011 (2012). Then, a co-bookmarking matrix for the remaining publications was created. On top of this matrix, I performed non-metric multidimensional scaling for ordination (using the R function *nmds*) and hierarchical clustering for detecting sub-areas within the papers (using the R function *hclust*). Afterwards, I employed a naming heuristic involving the APIs of Zemanta³ and OpenCalais⁴ to find cluster names.

Then, area size and location are determined. The center point of each circle was

³<http://zemanta.com>

⁴<http://opencalais.org>

calculated as the means of the coordinates of the publications based on the NMDS result. The size of the circle was determined by the number of combined bookmarks of the publications. Finally, a force-directed approach was chosen to unclutter the visualization and move documents into their respective areas. In D3.js, a force-directed layout for the documents was created. To simplify matters for collision detection, a circle was constructed around the documents. The centers of the areas were denoted as gravitational centers. Documents not within the limits of the area were instructed to move towards the gravitational center.

The visualization of the evolution of a conference over years is based on small multiples. It was first developed by Weiß ensteiner (2014). As far as time series visualization goes, there are many types of visualizations, most prominently index charts and stacked graphs. In the case of knowledge domain visualizations, simple visualizations are unfortunately not able to convey all necessary dimensions of the data (in terms of ordination, size of research areas and closeness). One possibility would be to use animation to show changes in a domain over time. Psychological studies have shown, however, that people are bad at recognizing change in an object or a scene. This phenomenon is called change blindness (Simons and Rensink, 2005). In small multiples (Tuft, 1990), a graph is drawn for each of the steps in a time series. This approach allows for direct visual comparison between different representations and is thus very relevant to knowledge domain visualizations.

2.3 Results

The result of the visualization procedure for a single year detailed above can be seen in Figure 5. The blue bubbles represent research areas. The size of the areas is determined by number of bookmarks that the papers have received. Spatial closeness implies topical similarities. As can be seen, “User modeling” is the area with most papers and most bookmarks. It is closely connected to several other larger areas, including “Recommender system”. A second cluster of areas can be found on the right hand side of the visualization, involving “Intelligent tutoring system”, “Adaptive system”, and “Problem solving”.

After the force-directed algorithm has finished, users can interact with the visualization. The interface was designed in such a way that most of the exploration can be done within a single window. The dropdown on the right displays the same data in list form. Once a user clicks on a bubble, he or she is presented with relevant documents for that area (see Figure 6). By clicking on one of the documents, a user can access all meta data for that document (see Figure 6). If a preview is available, it can be retrieved by clicking on the thumbnail in the meta data panel. By clicking on the white background, one can then zoom out and inspect another area. In addition, a user can filter the publications by entering terms in the search field on top of the list. Only publications that contain all of the search terms are

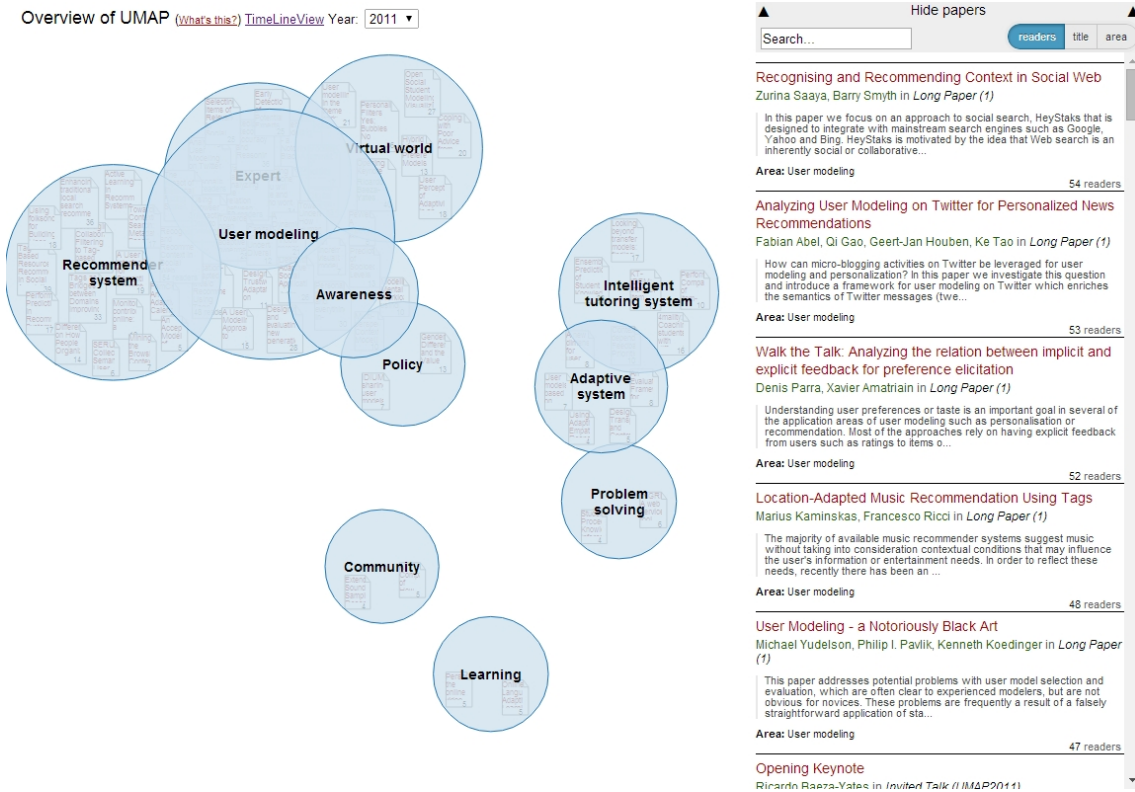


Figure 5: Topical visualization of UMAP 2011 based on co-bookmarking data.

displayed within the bubbles and the list. The list can also be sorted by title, area, and number of bookmarks to facilitate exploration via the list.

Figure 7 shows the evolution of the topical overview using small multiples. To aid the user in detecting changes between the representations, two visual helpers were introduced. First, a grid is drawn to help with comparing size and position of the research areas. Second, whenever users hover over an area, the corresponding area is highlighted in the other representation, and a line is drawn between the two entities. There are three areas that are present in both years: “User modelling”, “Recommender system” and “Intelligent tutoring system”. While the relative position of the areas to each other has not changed much, the area with the most papers and bookmarks is now “Recommender system”.

2.4 Discussion and Future Work

The first results from this type of analysis are encouraging. Using small multiples allows for a comparison of knowledge domain visualizations over various years. The work, however, also revealed several weaknesses of the current approach. First, the topology needs to be improved. As many of the areas are overlapping, it becomes

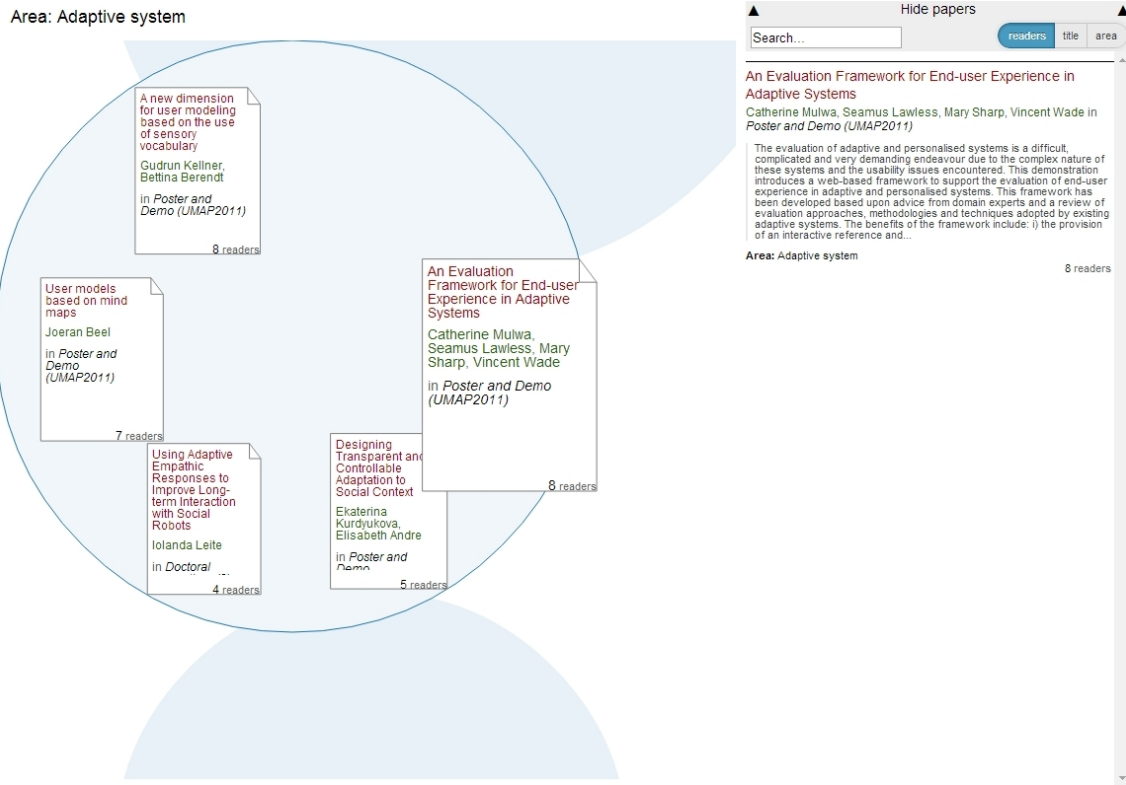


Figure 6: Zooming into the area “Adaptive Systems” with the paper “An Evaluation Framework for End-user Experience in Adaptive Systems” selected.

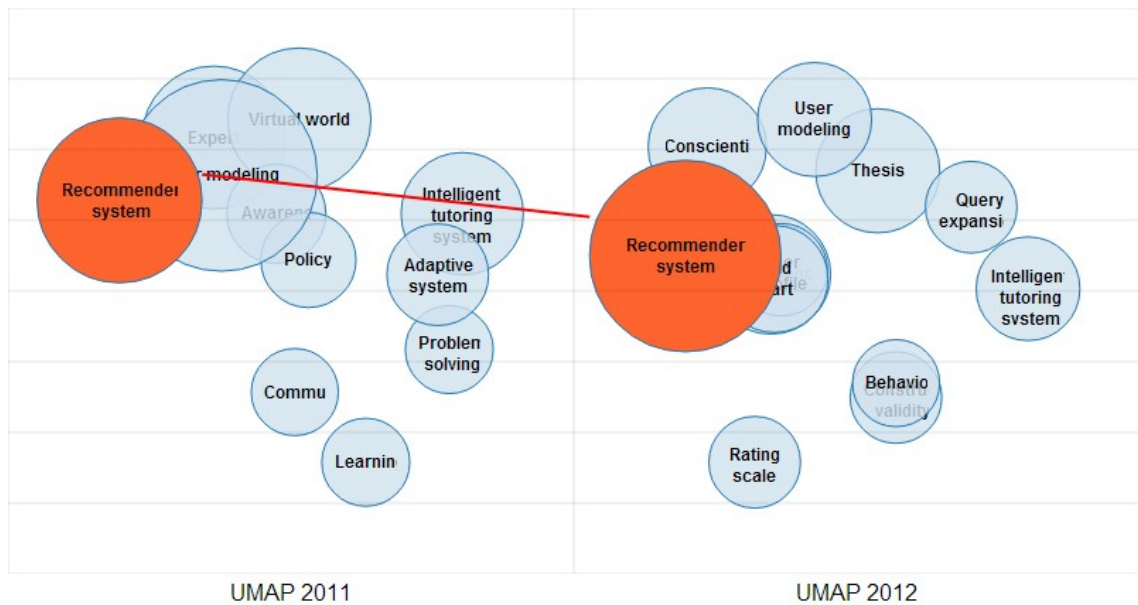


Figure 7: Evolution of the domain of UMAP from 2011 to 2012, with the area “Recommender system” highlighted

harder and harder to disambiguate between the different clusters. One way to overcome this problem might be to employ force-directed placement on the research areas.

Second, scheduling data in CN3 is sparser than readership data in Mendeley due to the fact that the audience of CN3 is restricted to conference participants. This means that the results of the clustering vary more when choosing different threshold values. It also influences the results which are based on the titles and abstracts in the clusters. Therefore, I want to explore supplementing bookmarks with content-based measures.

Third, the continuity between the two years is very low. There are only three areas that are present in both iterations. Therefore, it might be worthwhile to explore moving time windows of two years to show how the different papers move when associated with earlier or later years.

Finally, it will be important to evaluate the method and the interface. One way to go about this would be to ask experts to critically review the visualizations, and to give them the ability to manipulate size and location of papers and areas. This would allow for comparing the automatically created visualizations to the experts' perception of the field. Another possibility would be to contrast the visualization based on conference papers with the evolution of the field based on other types of literature (e.g. journal publications).

3 Adaptive Overview Visualizations based on Measures Derived from Article Content

3.1 Introduction

The visualization presented in section 2 is promising, but it also has an inherent weakness: being based on user interactions, it can only be created when enough bookmarking data is available. It therefore cannot be used before the conference begins - this, however, is the time when the system is most useful to conference participants.

One way of overcoming this weakness is to use content-based measures instead of bookmarking data. The downside of this approach is that research terminology has proven to be too fluent to provide consistent results over longer periods (Leydesdorff, 1997). Since we are dealing with a single year of a conference, however, this should not be an issue. As a first use case, the 12th Intelligent Tutoring Systems Conference (ITS 2014)⁵ was visualized.

3.2 Data and Method

The initial data was sourced from Conference Navigator 3. It consists of all papers in the conference with the following fields: title, abstract, type, and authors. For the ITS 2014 dataset, there are 32 long papers, 44 short papers, 39 posters, and 7 presentations in the young researchers track. All of the papers are in English. Then, cosine similarity was performed on titles and abstracts. This procedure can be seen in Figure 8.

At first the documents are retrieved from the system. Talks that do not have a paper are removed from the list, which lead to 122 papers. Then, the document metadata, consisting of titles and abstracts is retrieved. For this metadata, a certain amount of pre-processing is performed using the R package *tm*. First, the punctuation is removed (including intra-word dashes) and unnecessary whitespace is stripped. Then, the text is converted to lower case, and stop words are removed. Finally, the words are stemmed, to preserve only the word stem.

In a next step, the document-term matrix was produced with the R function *tm::TermDocumentMatrix*. On top of this matrix, cosine dissimilarity between the documents was calculated. The resulting dissimilarity matrix served as input to multidimensional scaling and clustering.

The results of multidimensional scaling and hierarchical clustering can be seen in Figure 9. In this figure, each symbol represents a document, and the different

⁵<http://its2014.its-conferences.com/>

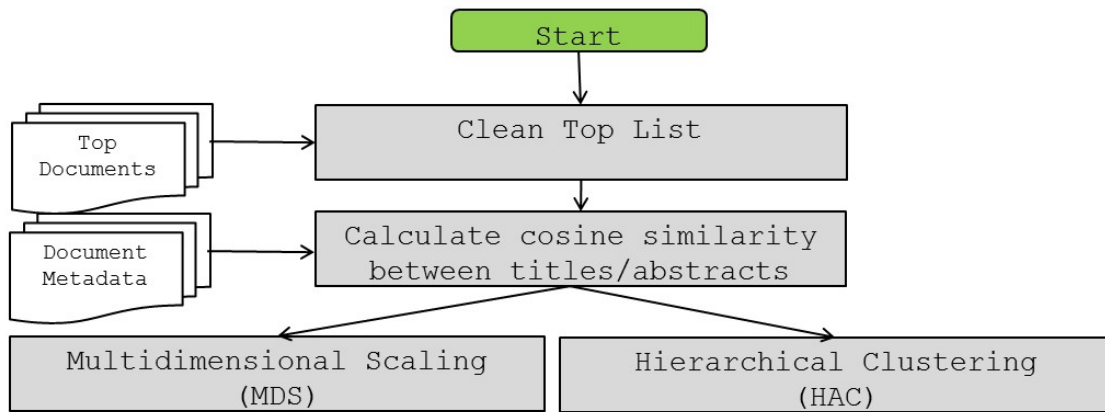


Figure 8: Visualization procedure employing cosine similarity

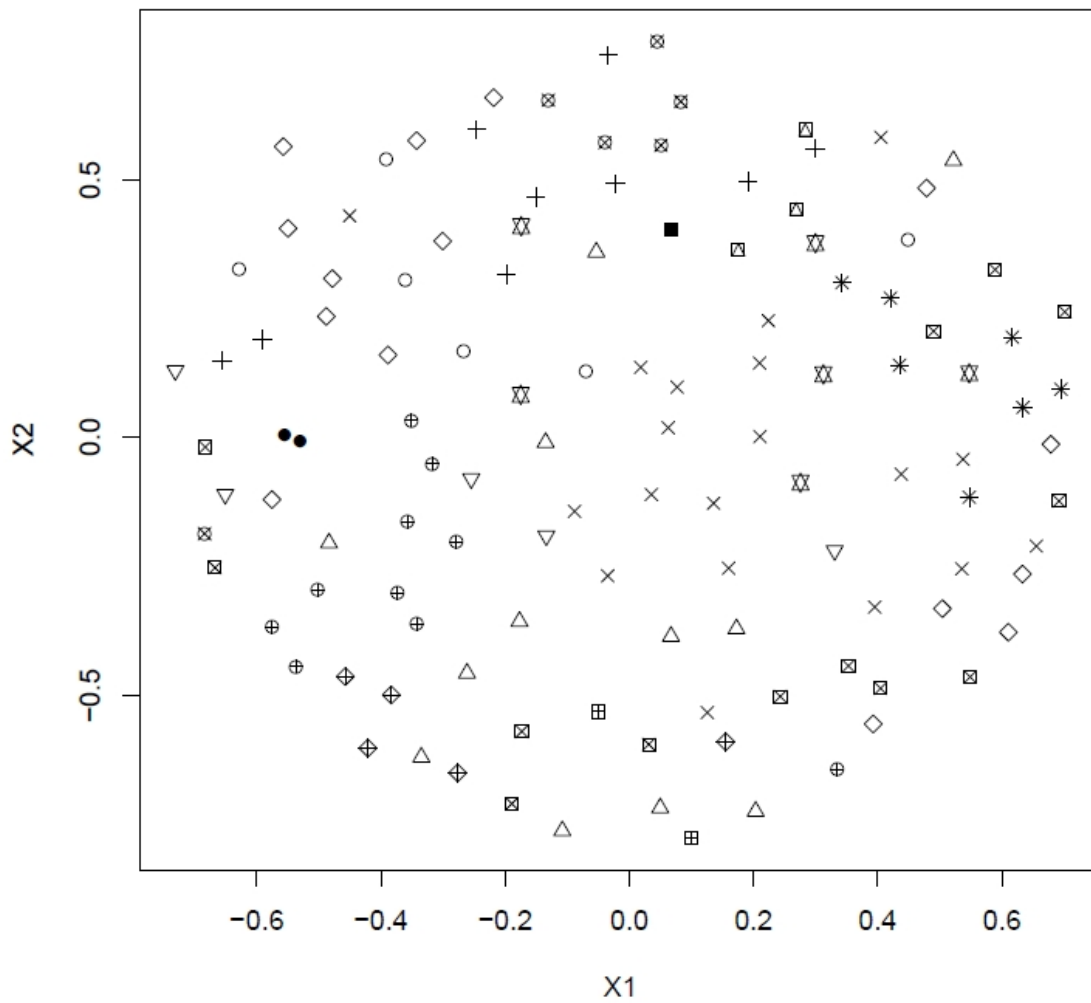


Figure 9: Results of hierarchical clustering and multidimensional scaling based on cosine similarity

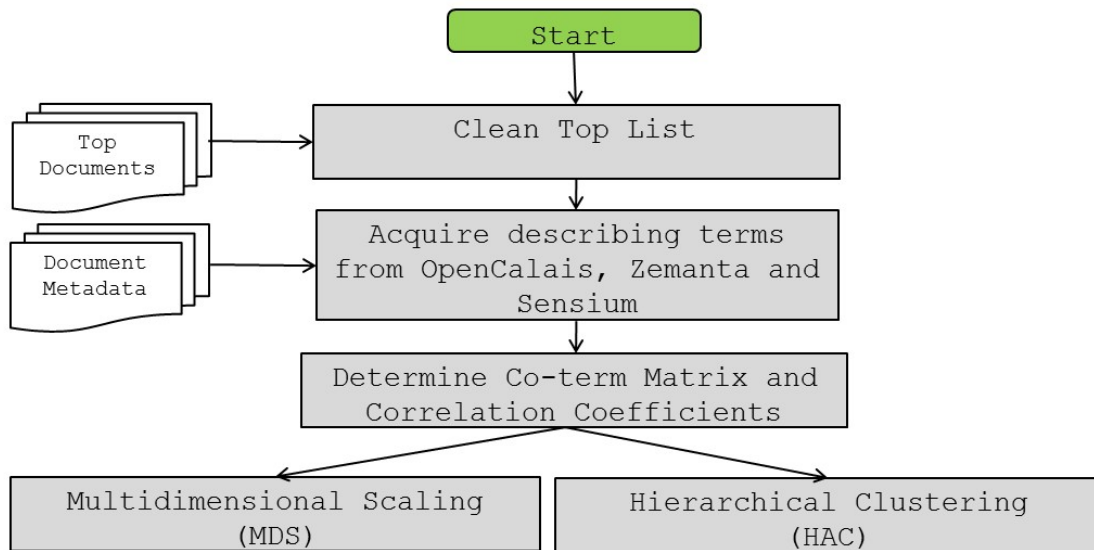


Figure 10: Visualization procedure employing co-occurrence of keywords

types of symbols represent different clusters. Unfortunately, indicators that judge the quality of outcome for both procedures suggest that the metadata was too dissimilar for processing. For MDS, stress was 0.29, and the R^2 was 0.41. In HAC, an elbow was found at 31% of explained variance. The visualized results confirm these indicators: the documents are uniformly distributed on the two-dimensional plane and the different clusters are not well separated from each other.

Due to the inferior performance on content from CN3 alone, external knowledge was acquired. Furthermore, instead of cosine similarity, a co-term approach was developed. The full procedure can be seen in Figure 10. The first step remains the same as above. Instead of using the document metadata directly, however, it was submitted to three different text mining services: OpenCalais, Zemanta, and Sensium. The first two services (OpenCalais and Zemanta) crawl the semantic web and return a number of concepts that the submitted text. Sensium extracts keyphrases from the submitted text. The titles and abstracts for each document were submitted to each of the three services. The resulting terms and concepts were stored in a database.

In a next step, a term co-occurrence matrix was calculated for all documents. Diagonal values were treated as missing values. The same goes for documents without term co-occurrences. Then, the correlation coefficients were determined. Finally, the Euclidian distances for the correlation coefficients were calculated, giving the dissimilarities between all possible pairs of documents.

This dissimilarity matrix was, again, the input to multidimensional scaling and hierarchical clustering. The indicators were more convincing this time (MDS: Stress=0.13,

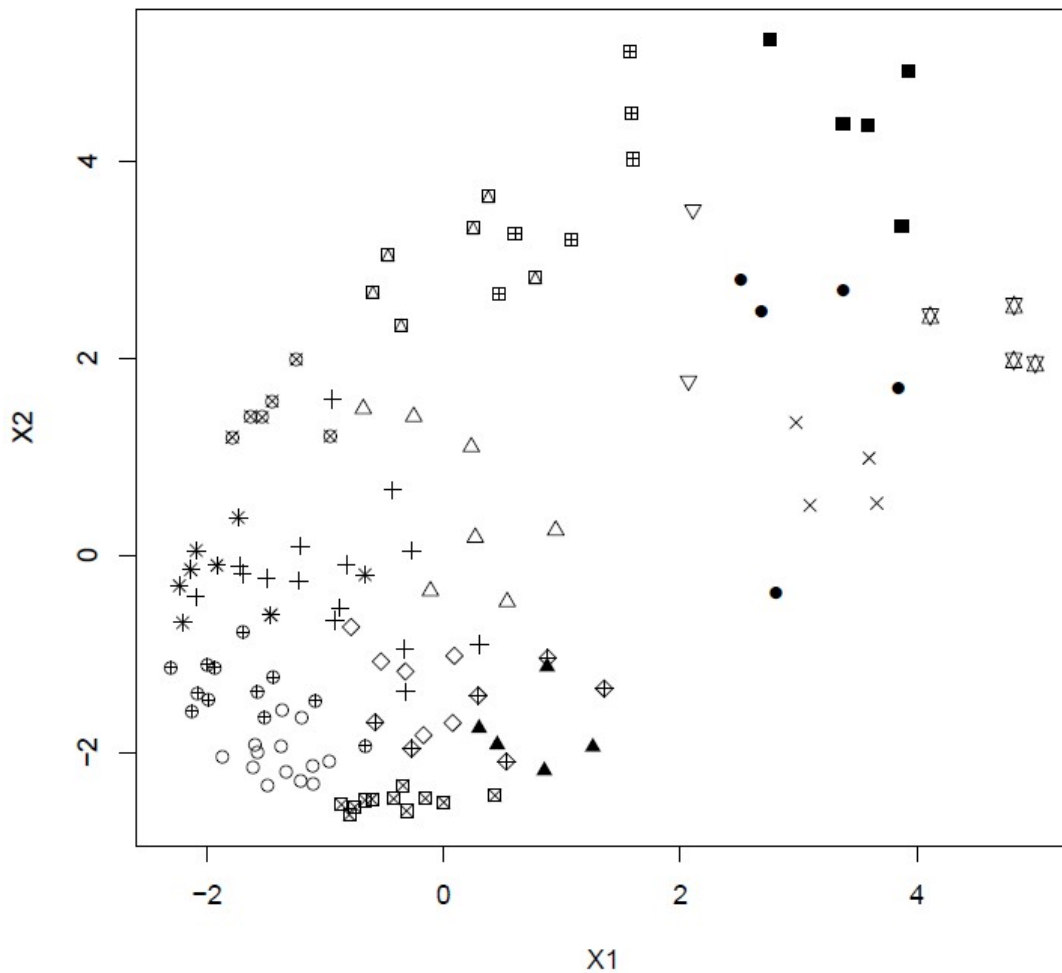


Figure 11: Results of hierarchical clustering and multidimensional scaling based on co-term similarity

$R^2=0.95$). The number of clusters (17) was chosen after manual inspection of the hierarchical clustering tree. The results of multidimensional scaling and hierarchical clustering can be seen in Figure 11. The chart emphasizes the indicator results: as one can see, clusters can now be much better distinguished from each other than in Figure 9.

On this basis, further steps towards an adaptive visualization were taken. The full procedure can be seen in Figure 12. The above steps are subsumed in “Co-term analysis”. In a next step, the clusters were named by hand by inspecting the papers they contain. Then, the papers and areas are uncluttered using a force-directed placement algorithm. The procedure is the same as described in section 2.2 with the exception that in the latter case only papers were uncluttered. This led to a result that was often hard to interpret as areas were overlaying each other.

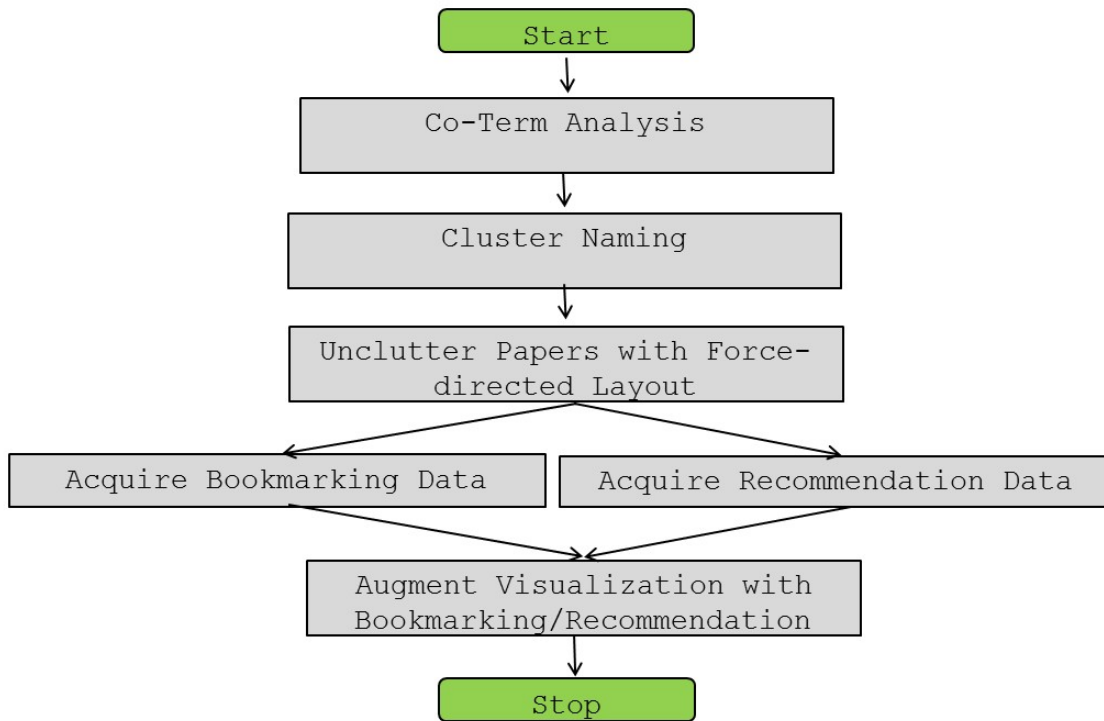


Figure 12: Full procedure of the co-term visualization approach

To allow for personalization of the visualization, the web application has been enhanced by a server component. So far, Head Start was a pure client-side application that was solely executed in the browser. A number of web services were added that talk to the database of CN3. They are REST-ful services written in PHP. JSONP is used as the exchange format. The following services have been implemented:

1. AddBookmark for adding bookmarks to the database
2. RemoveBookmark for removing bookmarks from the database
3. GetBookmarks to retrieve bookmarked and recommended items

In addition, an event tracking system for evaluation purposes was implemented. It logs user interactions with the visualization and the context in which the interaction was performed.

3.3 Results

The procedure detailed in the above section yields an overview of the ITS 2014 conference. The visualization contains 122 presentations in 17 areas. It can be seen in Figure 13.

When a user that is signed in to CN3 starts the visualization, it is adapted to the individual user profile. The interface for the adaptive version of the visualization sports the following new features:

1. All presentations that a user has already bookmarked are highlighted in green.
2. All presentations that are recommended for a user are highlighted in yellow. They are also specially marked in the list view with “Recommended”.
3. Presentations can be bookmarked right out of the user interface, by clicking on the appropriate button next to the title in the list view. In the same manner, talks can be un-bookmarked.
4. Papers are augmented with the number of bookmarks. This number is updated from the database each time the visualization is executed.

These features can be seen in Figure 14. Here, two papers have already been bookmarked, signified by the green border in the visualization. A third paper “Survival Analysis on Duration Data in Intelligent Tutors” is recommended for this user. This is also reflected in the list to the right, giving the user a way to bookmark or un-bookmark presentations.

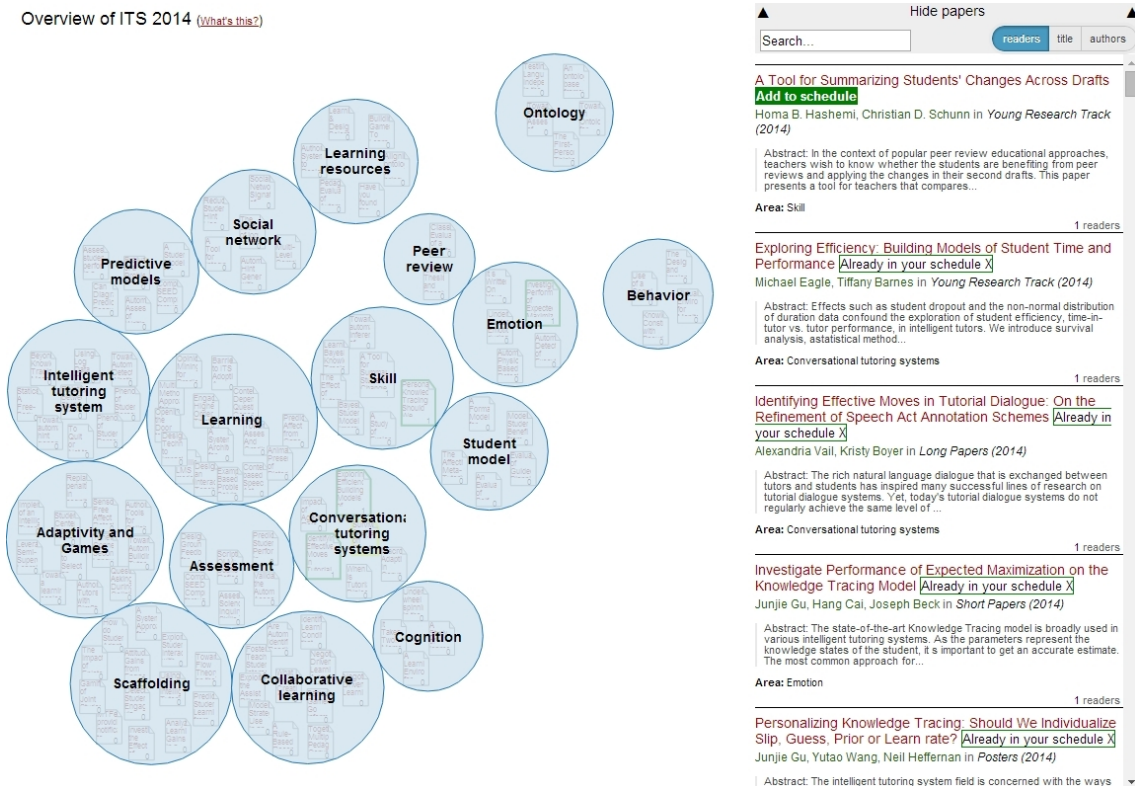


Figure 13: Overview of the Intelligent Tutoring Systems Conference 2014

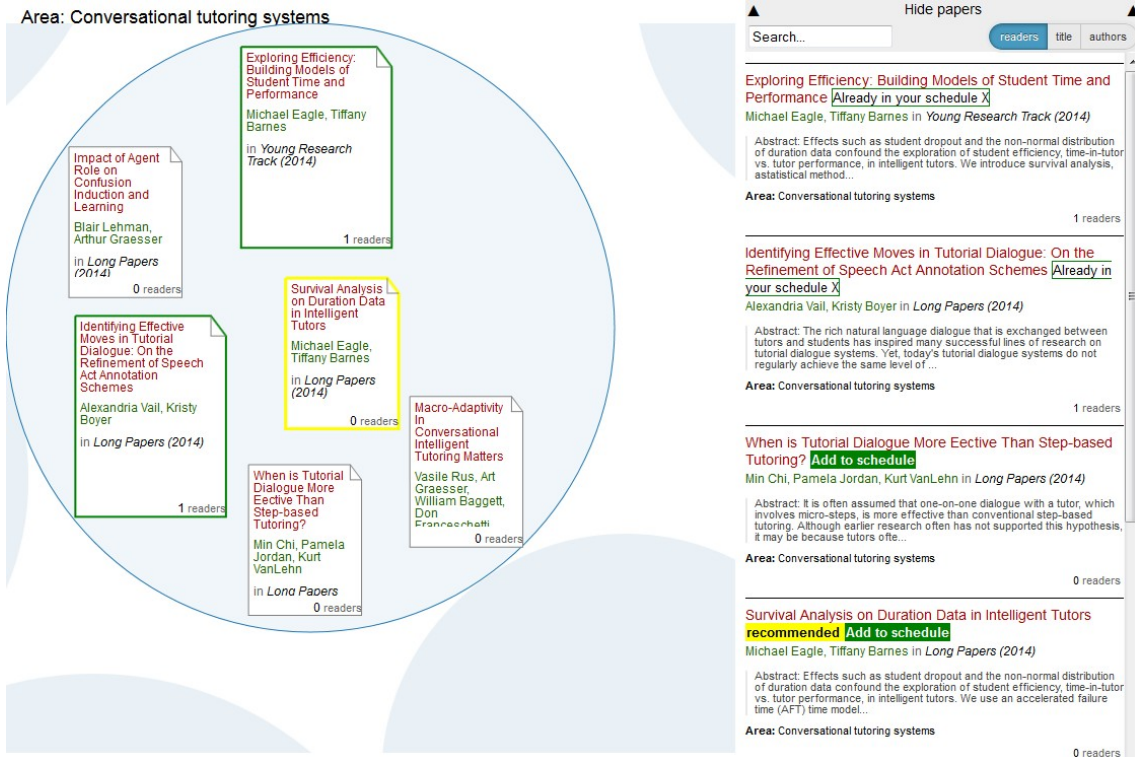


Figure 14: Zooming into the area “Conversational Tutoring Systems” with personalization features.

3.4 Evaluation Plan

In this section, an evaluation plan for the adaptive overview visualization is presented. This evaluation will be carried out in the context one of the upcoming conferences that Conference Navigator 3 supports. In this evaluation, both personalization features and usability aspects shall be covered.

In the evaluation, two hypothesis shall be tested:

Hypothesis 1: The more personalized clues users get, the more efficient they will be in bookmarking talks.

Hypothesis 2: The better users rate the usability of the visualization, the more efficient they will be in bookmarking talks.

These two hypotheses will be tested with a user study. In this study, conference participants will be asked to bookmark talks using the adaptive interface described in this section.

The evaluation will be carried out in four steps. In the first step, participants will be asked to provide demographic information, including age, profession, expertise and

their familiarity with visualizations. In the second step, participants are required to bookmark a certain number of talks using the conference schedule. In a third step, participants will be presented with the visualization and asked to bookmark further talks of interest. For this exercise, participants will be divided into four groups:

Group 1: Working with the full interface showing both bookmarked talks and recommended talks.

Group 2: Working with an interface showing only bookmarked talks.

Group 3: Working with an interface showing only recommended talks.

Group 4: Working with an interface with no personalization features.

In a fourth step, participants will be asked to fill out a brief survey regarding personalization features and usability. For personalization features, participants will be asked to rate the following statements on a 7-part Likert scale:

- I felt that I had enough information to find relevant talks.
- The clues presented by the system were useful to find relevant talks.

For usability, the System Usability Scale (Brooke, 1996) will be used with three additional open questions:

1. What did you like about the visualization?
2. What didn't you like about the visualization?
3. Do you have any suggestions regarding the visualization?

The results will be processed using three different measures. The first two measures will be derived from the log files collected during the participants' interaction with the visualization. Specifically, both the number of clicks before an item was bookmarked will be examined and the time needed during that process. Finally, the results of the Likert scale will be taken into account. Regarding the usability, the results of the enhanced System Usability Scale will be used.

4 Conclusions and Future Work

In section 1, the following research questions were posted:

RQ 1: What are adequate indicators derived from Conference Navigator for creating knowledge domain visualizations?

RQ 2: What are relevant personalization features for overview visualizations of conferences?

RQ 3: How to evaluate the adaptive knowledge domain visualization of a scientific conference?

With regards to RQ1, it could be shown that both bookmarking data and content-based measures can be used to create knowledge domain visualizations of conferences. The main drawback of bookmarking data is that it is only available when enough talks have been bookmarked. Content-based measures do not have this drawback. It, however, emerged that external knowledge is required to create a meaningful visualization; it cannot be created based on titles and abstracts alone as they proved to be too dissimilar to each other.

With regards to RQ2 and RQ3, the visualization was enhanced with highlighting bookmarked talks and recommended talks. Furthermore, users were enabled to bookmark talks directly from the interface. These features will be tested in a user study in which conference participants will be asked to bookmark talks using the adaptive interface. Time, number of clicks and a survey will be used to determine the efficiency of participants.

The work brought about many avenues for future research. For one the evolution of time-series data seems intriguing, but first results show that the continuity between years is very low. Therefore, it might be worthwhile to explore moving time windows of two years to show how the different papers move when associated with earlier or later years. In this regard, it would also be interesting to model the transition from a content-based to a community-based visualization (i.e. based on bookmarking data).

This would require that the creation of the visualization is changed from an offline to an online procedure. To reach this goal, several computational challenges have to be met, especially how to calculate the co-bookmarking patterns in real-time. A practical solution could be to pre-calculate the matrix and to perform only the creation of the map on the fly.

With regards to the interface, it would be interesting to include more publications in the visualization e.g. to show all years of a conference at once. This should be done without overwhelming the user. Thus it might be necessary to create a topical hierarchy within the clusters.

Finally, to achieve further adaptivity of the interface, it seems worthwhile to let users change the structure of the visualization in order that it corresponds better to their mental model of the conference. This would allow users to change the size and the configuration of areas, and the distribution of papers.

List of Figures

1	Growth of number of scientific journals on a logarithmic scale starting from the 17th century	2
2	Overview visualization of educational technology based on co-readership	3
3	Interface of Conference Navigator 3 showing the schedule of UMAP 2014	4
4	Knowledge domain visualization process	8
5	Topical visualization of UMAP 2011 based on co-bookmarking data. .	10
6	Zooming into the area “Adaptive Systems” with the paper “An Evaluation Framework for End-user Experience in Adaptive Systems” selected.	11
7	Evolution of the domain of UMAP from 2011 to 2012, with the area “Recommender system” highlighted	11
8	Visualization procedure employing cosine similarity	14
9	Results of hierarchical clustering and multidimensional scaling based on cosine similarity	14
10	Visualization procedure employing co-occurrence of keywords	15
11	Results of hierarchical clustering and multidimensional scaling based on co-term similarity	16
12	Full procedure of the co-term visualization approach	17
13	Overview of the Intelligent Tutoring Systems Conference 2014	18
14	Zooming into the area “Conversational Tutoring Systems” with personalization features.	19

References

- Amin, M. and Mabe, M. (2000). Impact factors: use and abuse, *Perspectives in Publishing* **1**(2000): 1–6.
URL: <http://www.ncbi.nlm.nih.gov/pubmed/14518149>
- Bollen, J. and van de Sompel, H. (2006). Mapping the structure of science through usage, *Scientometrics* **69**(2): 227–258.
URL: <http://www.springerlink.com/index/10.1007/s11192-006-0151-8>
- Börner, K., Chen, C. and Boyack, K. W. (2003). Visualizing knowledge domains, *Annual Review of Information Science and Technology* **37**(1): 179–255.
URL: <http://doi.wiley.com/10.1002/aris.1440370106>
- Brooke, J. (1996). SUS - A quick and dirty usability scale, *Usability evaluation in industry*, Taylor & Francis, pp. 189–194.
- Brusilovsky, P., Parra, D., Sahebi, S. and Wongchokprasitti (2010). Collaborative information finding in smaller communities: The case of research talks, *6th International Conference on Collaborative Computing: Networking, Applications and Worksharing (CollaborateCom)*, Vol. 15260, pp. 1–10.
URL: http://ieeexplore.ieee.org/xpls/abs_all.jsp?arnumber=5767048
- Farzan, R. and Brusilovsky, P. (2007). Community-based conference navigator, *Proceedings of 1st Workshop on Adaptation and Personalisation in Social Systems: Groups, Teams, Communities at the 11th International Conference on User Modeling*, pp. 30–39.
URL: <http://www2.sis.pitt.edu/peterb/papers/FarzanBrusilovsky.pdf>
- Garfield, E., Sher, I. and Torpie, R. (1964). The use of citation data in writing the history of science, *Technical Report 64*, The Institute for Scientific Information.
- Haustein, S. and Siebenlist, T. (2011). Applying social bookmarking data to evaluate journal usage, *Journal of Informetrics* **5**(3): 446–457.
- Hsiao, I. and Brusilovsky, P. (2012). Motivational social visualizations for personalized e-learning, *Proceedings of the EC-TEL 2012*, pp. 153–165.
URL: http://link.springer.com/chapter/10.1007/978-3-642-33263-0_13
- Kraker, P. (2013). *Visualizing Research Fields based on Scholarly Communication on the Web*, Dissertation, University of Graz.
URL: <http://media.obvsg.at/p-AC11312305-2001>
- Kraker, P., Jack, K., Schlögl, C., Trattner, C. and Lindstaedt, S. (2013). Head Start: Improving Academic Literature Search with Overview Visualizations based on Readership Statistics, *Web Science 2013*.

- URL:** http://know-center.tugraz.at/download_extern/papers/websci-cam-ready.pdf
- Kraker, P., Körner, C., Jack, K. and Granitzer, M. (2012). Harnessing User Library Statistics for Research Evaluation and Knowledge Domain Visualization, *Proceedings of the 21st International Conference Companion on World Wide Web*, ACM, Lyon, pp. 1017–1024.
- URL:** http://know-center.tugraz.at/download_extern/papers/user_library_statistics.pdf
- Larsen, P. O. and von Ins, M. (2010). The rate of growth in scientific publication and the decline in coverage provided by Science Citation Index., *Scientometrics* **84**(3): 575–603.
- URL:** <http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=2909426>
- Leydesdorff, L. (1997). Why words and co-words cannot map the development of the sciences, *Journal of the American Society for Information Science* **48**(5): 418–427.
- URL:** <http://www.leydesdorff.net/cowords/>
- National Science Board (2010). *Science and Engineering Labor Force*, Vol. 22 Suppl 1, National Science Foundation, Arlington.
- Parra, D., Jeng, W., Brusilovsky, P., López, C. and Sahebi, S. (2012). Conference Navigator 3: An online social conference support system., *UMAP Workshops*, pp. 1–4.
- URL:** http://ceur-ws.org/Vol-872/umap2012-poster_07.pdf
- Price, D. D. S. (1961). *Science since Babylon*, Yale University Press.
- URL:** http://pdfserv.aip.org/PHTOAD/vol_14/iss_6/46-1.pdf
- Price, D. J. D. S. (1963). *Little science, big science*, Columbia Univ. Press.
- URL:** <http://www.slac.stanford.edu/spires/find/books/www?key=8338>
- Simons, D. J. and Rensink, R. a. (2005). Change blindness: past, present, and future., *Trends in cognitive sciences* **9**(1): 16–20.
- URL:** <http://www.ncbi.nlm.nih.gov/pubmed/15639436>
- Small, H. (1999). Visualizing science by citation mapping, *Journal of the American Society for Information Science* **50**(9): 799–813.
- Tufte, E. R. (1990). *Envisioning Information*, Graphics Press.
- Weiß ensteiner, P. (2014). *Entwicklung einer Visualisierung für Time Series Data*, Bachelor’s thesis, Graz University of Technology.
- URL:** <https://github.com/pkraker/Headstart/tree/master/doc>

- Wongchokprasitti, C., Brusilovsky, P. and Parra, D. (2010). Conference Navigator 2.0: Community-Based Recommendation for Academic Conferences, *Workshop SRS'10*, ACM, pp. 1–5.