ABSTRACT
This paper describes the design, development and testing of a visualization tool targeted on publications, specifically on their keywords and authors — and the institutions they belong to.

Every year, billions of dollars in research funding are allocated to universities and the private sector by various investors. Each investor is usually interested in funding a different area of research, and their ultimate goal is to maximize the effect of that investment by selecting the best qualified organizations. Investors need to find organizations that maximize the productivity of research, for they ensure that all funding would be well spent.

This tool offers sponsors and research funding institutions, visualizations aimed to describe trends, activity volume, and inter-institutional work. Being able to observe this information in an organized manner will potentially enhance the user’s experience when analyzing data. Consequently, this tool will help the decision making process when assigning funds to research institutes.

It is worth mentioning that very few similar applications have been developed. Our aim is to contribute with a solution to this problem so that future research funding assigning investigation can be properly conducted.

Categories and Subject Descriptors
Software Engineering [Information Visualization]: Web Application

Keywords
infovis, research analysis, funding, trending topics

1. INTRODUCTION
We are living in the information era. Our desire to continuously expand our knowledge in all areas is reflected on the amount of time we dedicate to data and information related activities and the overall importance information has in our lives. Entities, such as universities, laboratories and research institutes, play a key role in the discovery of new technologies, development of strategies and adaptation of concepts into new scenarios. Consequently, they have become the primary beneficiaries of research targeted funding.

Technology transfer and work on recently discovered fields are important indicators of the activity of universities and research institutions. However, an accurate representation of these parameters is somewhat a convoluted task. Given the vast amount of universities and laboratories, it is often difficult to assign the adequate amount of funding to the institutions that require it. Tools and techniques to extract relevant information to develop a criteria on funding assignment are very scarce. It is not difficult to understand that the solution to this problem involves the participation of several data parameters, which can often be retrieved from research paper publications.

When performing an investigation on where to assign funding, the researcher must be able to tailor his search so that his information interests are satisfied. Insights on trending research topics, activity levels of universities in particular areas, and inter-institutional collaboration are cornerstones on behavior analysis to develop a funding assignment criteria. Research papers are able to provide all of the required data when they are published in standard formats.

For this reason, we have designed and developed an application to provide research funding assignment institutions with insights on information processed by the analysis of publications, presenting interactive visualizations that display information both on an overview and filtered manner, being capable to display details on demand.

It is worth mentioning that the value of these results is also potentially useful to research institutions. This application could assist the evaluation of current research plans and design of new, more flexible strategies. Study of these results could determine how to shape their funding gathering strate-
gies, so that more funding is obtained in the future.

A more informed decision criteria on research behaviors will undoubtedly better distribute research stipends. The products of this work favor both the sponsor and the research institution, since the two are able to determine with great accuracy the entities of their corresponding interest.

2. RELATED WORK

A popular alternative lies in techniques involving mathematical transformations, such as the one described by Daeh et al. [6], which uses linear regression analysis to show the correlation between the funding and research outputs.

The work of Wang et al in [17] analyzes the relationship between patents, research, and development spending using a finite mixed Poisson regression model with covariates in both Poisson rates and mixing probabilities.

Although these approaches are useful, we believe that a visualization generating tool is more appealing and would ultimately attract a larger amount of users. Card et al in [5] could not have stated it better by declaring that "human intelligence is highly flexible and adaptive, superb at inventing procedures and objects that overcome its own limits. The real powers come from devising external aids that enhance cognitive abilities... An important class of external aids that make us smart are graphical inventions of all sorts."

Nowadays, a wide variety of information visualization tools are available. We have particular interest in the analysis of the behavior of research publications and their related components, such as authors, institutions, etc. Most of the revised publications focus on providing academic literature tools and studies for research institutions. Bergstrom and Atkinson implemented PaperCube [3], that was designed to augment a scholar’s interaction with a digital library and explore bibliographic meta data using a defined set of visualizations.

Our approach is different. Given that our primary target are research funding entities, this tool shows users where the bulk of research papers are coming from year-to-year, providing a quick interpretation of which research institutions are more productive than others; i.e., this information establishes which research institutions produce a better return on investment. The possibility of visualizing the network of research collaboration shows where knowledge is transferred and how different research topics grow and shrink in popularity or have cyclic properties throughout time, therefore giving funding institutions a more detailed summary, so that resources can be allocated in several places and have an increased certainty of success.

Nevertheless, many of these papers explain interesting strategies and technology combinations for visualization production on academic literature studies.

Huang [8] proposed a new approach for collecting, analyzing and visualizing co-authoring data of individuals. This approach can be used for understanding the academic collaboration and knowledge domain of individual researchers in a past period, through repetitive co-published works. The InterRing visualizer shows the weight of co-authorship of an individual with other researchers in a particular academic year, as well as the knowledge domain of the individual that was covered by their previously published papers.

Closely related, Morris and Yen [13] propose a crossmapping technique that affords a method for exploring relations in a collection of papers. They provide examples of two types of crossmaps that show relations between research fronts, reference groups and author collaboration groups. Their technique allows the visualization of relationships between groups. It also permits group membership overlapping. Using this method, it is possible to visualize and understand the set of complex relationships among different groups, manifested in a knowledge domain.

Previous tools, as the one described by Lee et al in [10], were created to reveal connections, trends, and activity in the Information Visualization conference community. The tool, PaperLens, includes capturing the evolution of topics, giving users easy access to publications/authors, finding most frequently published authors, visualizing relationships between authors, and finding the most frequently referenced papers and authors.

On the other hand, Bertini and Lalanne [4] categorize visualization techniques in classes, highlighting current trends, gaps, and potential future directions for research. This is a particularly interesting discovery, since it embraces the concept of joint knowledge discovery using human and machine capabilities by integrating visualization and data mining techniques.

Action Science Explorer (ASE), described in Gove et al. [7], is a particularly innovative project. Although its goals are somewhat different to ours, it offers implementations of many of the concepts and ideas we consider vital to our application’s success. We find its method of rapid summary generation of academic literature as a solid guideline of what to keep in mind when developing a tool for studying author collaboration and recurring themes. Furthermore, it defines with great accuracy how to obtain visualizations that are capable of providing the user a clear understanding of data for consequent insight finding.

Few navigation options are enabled in the analyzed applications. A dedicated tool must be able to provide extraordinary visualizations, without disregarding ease of use. Multiple ways of input and navigation must be available so that versatile solutions to the same problem enrich the user’s experience while using the application. Having this in mind, we have explored different visualization techniques in which a large amount of data can be displayed without confusing the user.

FacetLens, as proposed by Lee et al. [11], enables users to identify trends and compare several of them simultaneously with an implementation of linear facets. It also offers pivot operations to allow users to navigate the faceted dataset using relationships between items.

TreeMaps, discussed by Johnson et al. [9], represent an attractive method for hierarchical representation of large amounts
of data. They offer a good way to represent information visually as a series of nested rectangles whose colors are determined by an additional measurement.

It is important to provide a unified fashion of displaying the visualizations, a characteristic some projects seem to fail to correctly implement. Biblioviz [15] presents an interesting approach on consolidation of desirable functionalities.

Most of the revised work include in their methods some variation of the visual information-seeking mantra, as originally proposed by Ben Shneiderman [16]. They have successfully provided a starting point. Our goal is to go further by showing which institutions are most active through time periods, which ones collaborate the most with each other, and also detect trends in research topics.

3. PROJECT DESCRIPTION
This application has been carefully designed under a minimalist approach. We include characteristics such as only one customizable screen, so users can focus all their attention to one part of the screen for control, therefore avoiding distractions or confusion.

A bustling discussion on the impacts and results of previous work was necessary to reach conclusions that later contributed to our tool’s design. We observed that some applications are operating system dependent or required a high amount of resources to work efficiently. Related material, such as [2] state the drawbacks of specific system implementations, which are only available for a restricted number of users.

These type of constraints greatly limit portability, which in our opinion must always be among the application’s primary characteristics. For this reason, this tool used a web application structure. Such architecture enables our product to reach a very large amount of the potential users with minimal extra requirements. We agree with Bajaj and Cutchin’s [1] conclusion that the use of a web application architecture provides great flexibility to the product.

Additionally, a web application implementation will always provide a relatively light visualization elaboration tool than can be easily deployed in almost any computer.

The server side programming was implemented in PHP and MySQL, being PHP a programming language and MySQL a database engine optimized for web applications. All visualizations were handled with D3, a JavaScript library which allows arbitrary data binding to a Document Object Model (DOM), and then apply data-driven transformations to the document.

D3 is not a traditional visualization framework. Rather than providing a monolithic system with all the features anyone may ever need, D3 solves only the crux of the problem: efficient manipulation of documents based on data. This gives D3 extraordinary flexibility, exposing the full capabilities of underlying technologies such as CSS3, HTML5 and SVG. It avoids learning a new intermediate proprietary representation. With minimal overhead, D3 is extremely fast, supporting large datasets and dynamic behaviors for interaction and animation.

Typical required adaptations, styling, and user enhancing experience implementations were handled with HTML, CSS and JavaScript.

We implemented a tool that generates real time visualizations with details on demand. Our application focused on providing to the user information to develop decision making criteria. We put special emphasis on:

**Institution Activity:** Displaying which institutions are the most active on publishing papers on a specific topic.

**Collaboration:** Visualizing the institutions that collaborate with each other to publish papers.

**Trend:** Determining how different research topics grow and shrink in popularity over time.

A treemap (Fig. 1) was included as a general navigator from which most of the available filtering can be performed without the use of a keyboard. We decided to implement one in order to take advantage of the treemap’s ability to easily display patterns that would be difficult to spot in other ways. Their efficient use of space was also vital in our selection, due to the usual display resolution constraints found in everyday computing. Furthermore, they supply ease to zoom in and filter the data by analyzed keywords, providing information-rich visualizations a with a reduced amount of clicks.

![Figure 1: Treemap Visualization](image)

Their ability to legibly display thousands of items on the screen simultaneously [9] made them the perfect general navigation visualization. By enabling clickable areas to zoom into a specific field, we were able to offer filtering and consequent details on demand. We found this particularly attractive, since its inclusion enhances our application to allow multiple ways of input. Ultimately, this addition allows the user to enjoy a more satisfying experience with our application.

We determined that stack charts, which show the relationship of parts to the whole, were to be used when displaying visualizations on keywords (Fig. 2), authors (Fig. 3) and institutions (Fig. 4). Each stack chart was assigned a general color to help the user remember and distinguish visualizations from one another. The colors we chose are:
• Blue, for the keywords stack chart.
• Magenta, for the authors stack chart.
• Green, for the institutions stack chart.

Stack charts have been proved effective when representing divisions within data label categories or stacking two separate numeric categories. They were able to display very effectively large amounts of data without loss of comprehension. Stacked charts offer similar complexity to clustered charts by adding together component value items within chart bars or areas. By stacking items and assigning a different color to each item, you can effectively display trends among comparable or related items, or visually emphasize a sum of several indicators.

A careful integration of the above described methods and visualization techniques allowed us to implement Shneiderman’s [16] “overview, zoom and filter, details-on-demand” in our system.

The dataset analyzed in this work was obtained from sampling the Transportation Research Board (TRB) journals from 2007 to 2010. The TRB is an institution dedicated to promote research on transportation solutions. This particular dataset includes standard information on publications, such as authors and the institution they belong to, as well as publishing date and keywords. It is worth mentioning that adequate practices have been implemented so that introducing a new dataset into the tool represents minimal work and effort. This measure was taken in order to guarantee our tool has sufficiently scalable and flexibility characteristics.

Curiously, we found a challenge in obtaining a well kept academic literature reference dataset. Many of the most popular sites providing this service do not offer free snippets of their database for academic purposes. When searching for free of costs datasets, we found a vast collection of outdated information displaying incoherent data or having missing fields. Because of this, all free datasets were discarded and decided to build our own using TRB journal’s entries.

Our application as been hosted in a virtual server machine with minimal resources, since most of the processing is done locally. This application works in Internet Explorer 9, Firefox 3.6, Chrome 11, and Safari 5, and later versions, enabling it to reach the vast majority of users.

The visualizations have been optimized for display screens with a 1680 x 1050 resolution. However, visualizations can be resized, so users can adapt them to their screen resolution. Fig. 5 (available as the last page of this paper) depicts a screenshot of our fully operational application.

3.1 Evaluation
Using the interface that we developed at this point, we conducted a usability test using a traditional user interface evaluation technique, called think-aloud [14]. Thinking aloud asks participants of usability test to literally say as much as what they are thinking and is proven to be effective in many...
user interface design process. Since our purpose was not to measure the performance of how fast users complete tasks, we emphasized that we are not recording their performance to users. So users will not focus too much on the task, but they can relax and express what they are thinking. We also asked users to fill in a survey to grade overall usability of the interface.

Although our primary target users are professional investors and researchers who needs to know about unfamiliar research topics, because such experts work for days and weeks to carry out their actual tasks, it is very hard to ask them to participate in the usability study. Therefore, we decided to ask people who have research experience to ensure that participants have domain knowledge. We also asked if participants are fluent in using computer in general, web browsers, and other visual analytic tools, such as Spotfire and Tableau. Participants were all familiar with how to use computer and browsers, and most of them had strong experience in using visual analytic tools. These facts suggest that participants have sufficient similarity in characteristics to our intended users.

For the evaluation, we explained participants the intended users and tasks of the interface, and explained the functionalities of Research Journal Network. For the test, we used a small monitor (13 in) and a big monitor (15 in) to see the effect of screen size to usability. We asked participants to complete following five tasks to see how they interact with the interface.

### 3.1.1 Task A
We asked to open a “keyword chart”, the stack graph which describes temporal change in number of publication under certain keywords. Participants were asked to select a keyword traffic and describe the change in number of publication between 2007-2008, 2008-2009, and 2009-2010.

We noticed that after participants open the keyword chart, half of the participants went straight to search box or pull down menu on the control panel to accomplish the task users went. We were interested because we expected users to use the keyword chart to search and filter the keyword.

Once users selected the keyword traffic, five out of six users succeeded in answering the questions asked on the task. From this, we became confident that the interface provides an appropriate view to show the change in number of publication over years. However, we also realized that there are several shortcomings of the interface. For example, there was a conversation as follows:

**Participant:** “I’m looking for... Is there undo?”

**Examiner:** “Yes, there is a small text down there saying Undo.”

**Participant:** “Um... Oh.”

This shows that the keyword chart and other charts need to provide clearer undo button.

### 3.1.2 Task B
Task was to open the “institution chart” that describes temporal change in number of research papers published by different institutions. Then participants were asked to determine the institution with the most publication throughout the years shown. For this task, one of the participant expressed a confusion as follows:

**Participant:** “Total? Is this... (pointing 2007).”

**Examiner:** “No, that’s 2007. (There was no label for 2007 on time-axis).”

**Participant:** “Is there a way to see total?”

**Examiner:** “No... We will implement it.”

**Participant:** “Ok. Hnn, to me, it looks like the first one is publishing the most. Bar height in 2007 is the highest, 2008 is the highest, 2009 is the highest, 2010..., probably not, but still this looks like the highest.”

This conversation suggests three insights. First, there is a design deficit of not showing the first year (2007) on the horizontal axis. Secondly, the interface does not provide the total number of publications of each institution, so participants had to solely depend on their intuition. Thirdly, even we did not provide the detailed number of publication, people were capable of finding out which institutions are publishing the most. The above conversation provides mental flow of what participants were thinking while accomplishing this task and we can see a stack chart a great way to represent the total number of publications.

### 3.1.3 Task C
Participants were asked to open the collaboration visualization from the control panel, then see what institutions are co-authoring with which institutions and how much. Because the visualization was not fully implemented, we used the Wizard of Oz technique [12], meaning we hard-coded the visualization for the evaluation purpose and asked users to accomplish the task. We chose North Carolina State University as the center institution, and defined five institutions that collaborate with the institution.

First, we asked participants to figure out which institutions the center institution is collaborating with. Most of the users clicked on icons that represent the collaborators and had no problem finding out the names of collaborators. However, for the second part of the task, users struggled to find out how to find numbers of collaborations between universities. For five out of six participants, examiner had to tell them that they can click on each edge to show the number of collaborations. Even after telling participants to click on edges, it was hard because edges were small.

On the post evaluation survey, this visualization was very controversial. Apart from the fact that it is hard to show details by clicking edges, some participants liked the visualization, because the visualization gives information in a simple way. On the other hand, some participants did not like the visualization because the aesthetically it does not
match with the rest of visualizations. Also, many participants did not like the fact that they have to close the balloon every time.

3.1.4 Task D
For this task, we asked participants to open the treemap from the control panel. Then asked to find a keyword Ship, and click to see detail underneath and find institutions publishing under this topic the most. Overall, the task seemed straightforward to participants. Some participants were struggling because after they click on the keyword, it shows institutions in small area. Even though the detailed information was shown on top-left corner of the chart, participants did not notice until the examiner told them about it.

3.1.5 Task E
On this task, we asked user to use the features on the control panel. Participants were asked to type in a keyword Maryland into the search box and check if the effect of filtering was clear. Every participant went straight to the search box within the control panel without hesitation. This shows the effective design of all the control at the left control panel.

3.1.6 Comments
Overall, we got following positive comments:

- Nice layouts
- Clean and simple control panel
- I liked the animations, because they were pretty
- Google map was straight forward
- The colors!

And negative comments as follows:

- No zoom available on treemap.
- When windows are closed, it is not clear different visualizations are interacting together

4. CONCLUSIONS AND FUTURE WORK
We have developed an application that offers an innovative and portable tool for research funding investment analysis, by providing insightful visualizations of behaviors in academic literature.

Many visualization techniques and methods were included and carefully combined to create a tool capable of displaying information on trending topics over time, as well as displaying collaborations among institutions and most active researchers.

The aggregated value of our work lies not only on the visualizations displayed, but also on the understanding of the concepts and relationships involved in a research process, so that the application maximizes important characteristics of the provided information. Consequently, this allows the user to perform quality insight extractions with relatively less effort and reasoning.

Future work is related on enhancing the user’s experience with the tool. Analyzed modifications include the presentation of the abstracts of filtered papers, search results saving capability, and a user friendly dataset import module.

5. ACKNOWLEDGMENTS
We would like to thank Dr. Ben Shneiderman for his valuable support and guidance throughout the whole process of making this project a reality.

6. REFERENCES


Figure 5: Application Screenshot