ABSTRACT
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.

Our tool, Research Journal Network (RJN), gives investors a detailed summary of research activities, so that resources can be allocated more effectively. 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. RJN 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.

Our general contribution relies on providing visualizations of large amounts of different research investigation data, so that funding agencies can better analyze their funding decisions. We implemented our tool in a way that the user can perform both high level abstractions and highly detailed analysis of specific research institutions/researchers in a short amount of clicks.

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 assign-
adequate amount of funding to the institutions that require it. Tools and techniques to extract relevant information to develop a criteria for 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 funding agency must be able to tailor their search so that their information interests are satisfied. Insights on trending research topics, activity levels of universities in particular areas, and inter-institutional collaboration are cornerstones of 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.

It is very difficult for human beings to effectively analyze and extract insights of an extensive set of numbers or values with little order or classification (i.e reading a spreadsheet with the top 1000 highest valued stocks behavior over the last decade). A visualization tool is ideal for displaying large amounts of related data with several parameters.

The rest of this paper is organized in a way that the reader finds this document complete and easy to read. The related work section describes relevant tools, methods, and techniques studied to build the knowledge base for RJN’s design and implementation. Project description presents the tool’s architecture, its design characteristics, and implementation approach. Evaluation narrates the testing methods used in order to quantify the application’s performance. Results and discussion exposes the results obtained by testing and develops solutions to the problems detected. Finally, conclusions and future work develops on RJN’s success and outlines improvements to come.

2. RELATED WORK

Previous attempts at modeling the performance of an institution that has received funding lies in techniques involving mathematical transformations, such as the one described by Dash et al. [6], which uses linear regression analysis to show the correlation between the funding and research outputs. The work of Wang et al. [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. This work is useful for obtaining a general idea of the effectiveness of research leading to an innovative result. However, we find their results too general and rigid for on demand analysis on a granular level.

Although these approaches are useful, this paper describes a visualization generating tool, given visualizations are effective in representing large amounts of data in an easy to understand way. Card et al. [5] stated 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.”

A wide variety of information visualization tools are available. This paper particularly focuses 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. Bergström and Atkinson implemented PaperCube [3], a visualization tool that was designed to augment a scholar’s interaction with a digital library and explore bibliographic meta data using a defined set of visualizations.

RJN’s approach is different than the ones cited above, 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. It is worth mentioning that we do not take grant amounts as input, but instead provide visualizations to determine research activity and collaboration based on the analysis performed over information obtained from publications.

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.

Many of the papers detailed below 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. [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.

Contrasting this, Bertini and Lalanne [4] categorize visualization techniques in classes, highlighting current trends, gaps, and potential future directions for research. This is a particularly interesting approach, 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 the ones of this project, it offers implementations of many of the concepts and ideas we consider vital to our application’s success. ASE offers a solid guideline of what to keep in mind when developing a tool for studying author collaboration and recurring themes by a rapid summary generation of academic literature. 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 explored the different visualization techniques in which a large amount of data can be displayed without confusing the user, which are described below.

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. This is effective when trying to visualize the correlation between two parameters in a dataset.

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.

Shneiderman [16] developed an abstraction capable, detail-on-demand visualization strategy that is very useful when displaying large amounts of data.

3. PROJECT DESCRIPTION

This application has been carefully designed under a minimalist approach. RJN adopts a single customizable screen approach, so users can focus all their attention to one part of the screen for control, therefore avoiding distractions or any confusion.

A discussion on the impacts and results of previous work was necessary to reach conclusions that later contributed to RJN’s design. Some papers describe applications that were operating system dependent or required a high amount of resources to work efficiently. Related material [2] states the drawbacks of specific system implementations, which are only available for a restricted number of users.

These type of constraints greatly limit portability, which RJN promotes as one of its primary characteristics. For this reason, this tool used a web application structure. Such architecture enables the application to reach a very large amount of the potential users with minimal extra requirements. RJN supports Bajaj and Cutchin’s conclusion [1] that the use of a web application architecture provides great flexibility to the product. Additionally, a web application implementation will always provide a relatively lighter visualization elaboration tool than can be easily deployed in almost any computer.

The server side programming was implemented in PHP and MySQL. Both PHP and MySQL are 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 datasets 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.

RJN is hosted in a virtual server machine with minimal resources, such as a 2 Gb hard drive, and 4 Gb of RAM, since most of the processing is done locally. The tool was tested to work in Internet Explorer 9, Firefox 3.6, Chrome 11, and Safari 5 by our development team. It is assumed that it will also be fully functional on later versions of these browsers. Nevertheless, its extended compatibility enables it to reach the vast majority of users. The visualization scales to the user’s screen size, so there is no need for the user to change or manipulate the size of the windows unless they choose to do so. Fig. 6 (available as the last page of this paper) depicts a screen shot of our fully operational application. A tool that generates real time visualizations with details on demand was implemented. This application focused on providing to the user information to develop decision making criteria. RJN 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.

3.1 Displaying the information

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. RJN has one implementation of the treemap technique in order to take advantage of its ability to easily display patterns that would be difficult to spot in other ways, such as the participation of a research institution on a particular topic compared to other institutions. 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 with a reduced amount of clicks.

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. This is 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.

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) and how they change over time. 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 proven effective when representing divisions within data label categories or stacking two separate numeric categories. They are 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, the user can effectively display trends among comparable or related items, or visually emphasize a sum of several indicators.

Finally, collaboration is depicted with a Google Maps implementation (Fig. 5). If the user clicks on an institution in the Institutions Visualization, the Google Map will be populated with that institution’s collaborators. Collaborators are other institutions that have co-published papers with the selected institution. The visualization is centered around the selected institution, and map markers are placed to indicate the collaborators. Clicking on a map marker opens an information window that displays the name of the institution and the total number of collaborations. The width of the lines connecting the selected institution to the map markers..
indicates the number of collaborations, making it easy for a user to spot a frequent collaborator.

Figure 5: Collaboration Visualization

An increase in the number of input fields will directly affect the learning curve for a new user, since they will probably read all of them before making a decision on where to input data. The tool’s interface is designed so that minimal input fields are included, yet it offers several ways of obtaining the same results so that a sense of integration is obtained while keeping granularity in the back end.

3.2 Dataset
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 is scalable and flexible.

Curiously, when searching for free datasets, we found that most of it was outdated, incoherent or incomplete. Many of the most popular sites providing this service do not offer free snippets of their database for academic purposes. When searching for free 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.

3.3 Evaluation
A usability test was conducted using a traditional user interface evaluation technique, Thinking Aloud [14], which is proved to be effective in user interface design processes. Thinking Aloud asks the participants of the usability test to literally say as much as they can about what they are thinking while testing the application. The tasks to be performed by the testers were designed in such a way that they would start from a rather simple task and escalate into a more convoluted one, making all of them part of an analysis routine.

Our primary target users are investors, funding agencies’ staff and researchers who needs to know about unfamiliar research topics. It is very difficult to obtain a group of people that meet these requirements because of their tight agenda. A solution to this problem was found in people who have research experience. By this way, we ensure that participants have domain knowledge. Our tester population consisted of six fellow students from our Information Visualization Class at the University of Maryland, at College Park. A brief description of our tester’s profile includes high computer proficiency, college instruction level, familiarity with visual analysis tools, such as Spotfire and Tableau, and an age range between 23 to 28.

The test’s purpose was not to measure the users speed to complete tasks. It was emphasized that they were not being recorded. This created the adequate atmosphere for users to relax and express what they are thinking instead of focusing too much on the task. The participants were shortly briefed on who the intended users are, the tasks to be done using the interface, and the functionalities RJN offers.

Two monitors, a 13 inch and a 15 inch were used to observe the effect of screen size in usability tests.

The testers completed the following five tasks:

3.3.1 Task A
The testers had to open a “keyword chart”, the stack graph which describes temporal change in the number of publications under certain keywords. Participants had to select the keyword traffic and describe the change in number of publications between 2007-2008, 2008-2009, and 2009-2010.

3.3.2 Task B
This task involved opening the “institution chart”, which describes the change in the number of research papers published by different institutions through time. The testers were asked to determine the institution with the most publications from 2007 to 2010.

3.3.3 Task C
Participants were asked to open the collaboration visualization from the control panel. Then, they had to determine which institutions were co-authoring papers and how many occurrences of this particular event existed. The Wizard of Oz technique [12] was used, meaning the visualization was hard-coded for evaluation purposes because it was not fully implemented. North Carolina State University was chosen as the center institution, and five other institutions were defined as collaborators.

3.3.4 Task D
For this task, we asked participants to open the treemap from the control panel. Then, they had to find the keyword Ship, and click on it to see its details underneath. Finally, they were required to find the most active institutions on this topic.

3.3.5 Task E
Testers were to use the features on the control panel for this task. Participants were asked to type in the keyword
Maryland into the search box and check if the effect of the filtering was clear.

Finally, users were asked to fill in a short survey to grade the overall usability of the interface. They had a score range between 1 (lowest) and 9 (highest). The survey was composed by the 5 questions described below:

- Application’s ability to show trending keywords over time.
- Application’s ability to show the institutions’ activities (number of publications).
- Application’s ability to show collaboration among institutions.
- Overall ease of use of the application.
- Overall performance of the application.

4. RESULTS AND DISCUSSION
For task A, five out of six users correctly answered the questions relevant to the task after selecting the traffic keyword. This provides confidence to the idea that RJN effectively displays the change in the number of publications over years. Half of the participants went directly to the search box or pull down menu on the control panel after opening the keyword chart to accomplish this task. This was an interesting behavior since we expected users to use the keyword chart to search and filter the keyword instead.

We also realized that there are several shortcomings in the interface. For example, we had a conversation similar to the one below:

Participant: “I’m looking for... Is there undo?”
Examiner: “Yes, there is a small text down there saying Undo.”
Participant: “Um... Oh.”

The first improvement for RJN in future work was then inferred. The keyword and all other charts need to provide a more evident undo button.

For task B, one of the participants expressed a confusion in a dialog similar to the one below:

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. Hmm, 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 short conversation provided us three insights about RJN’s design and implementation. First, there is a design flaw in not showing the first year of the time interval, 2007 in that particular case, on the horizontal axis. Second, because the interface does not provide the total number of publications for each institution, the user is left to depend solely on their intuition. Finally, users are capable of finding out which institutions are the most active researchers even if an exact number of publications per institution is not provided. A stack chart would a great way to represent the total number of publications in future work.

In task C, most of the users clicked on icons that represented collaborators and had no problem finding out their names. However, users struggled to find out the amount of collaborations between universities. Five out of the six participants asked for help. The examiner had to tell them that RJN allow clicking on each edge to show the number of collaborations. In their feedback, they commented that the edges were too small. Evidently, an improvement was required in this matter, so the edge width was increased, and the number of collaborations was moved to the institutions’ icons, instead of the edges.

This visualization was very controversial. Even though it is hard to display the details by clicking edges, some participants liked the visualization because it provides information in a simple way. On the other hand, some other testers did not like the visualization because it aesthetically does not match with the rest of the visualizations.

Task D seemed straightforward to all participants. Some participants struggled after they clicked on the keyword, since it shows all the relevant institutions in a small area. The detailed information was displayed on the top-left corner of the chart. However, participants did not notice it at all. Implementing a way to display this information in a more noticeable manner is the consequent solution.

As for task E, all participants 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.

RJN was graded pleasantly enough, averaging 7.5 on all of the questions in the survey. A few positive comments we made mostly on RJN’s layouts, clean and simple control panel, colorful animations, straightforward use of Google Maps, and the overall use of colors. Some negative comments, such as no zooming available for the treemap, and that it is not clear when different visualizations are interacting with each other were also received.

5. 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 the dataset’s potential to provide valuable information to the investors. Consequently, this allows the user to perform quality insight extractions with relatively less effort and reasoning.

Given the usability evaluation result, we fixed interface’s user navigation components. On the charts with temporal data, it now shows detailed information on the information panel. Collaboration visualization now shows number of collaboration when users click on institution icons, instead of clicking on edges. Although these are small changes, we believe improving such interface navigation improve user experience very much. In future work we plan to further enhance the user experience. Many of them are described in the discussion section. After some later examination, modifications including the availability of the abstracts of filtered papers, the ability to save search results, and a user friendly dataset import module will be implemented.

6. 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, and Michael Pack for sponsoring the project.

7. COLLABORATION
RJN was designed and implemented by this paper’s authors. All of them contributed the same amount of work and effort to make this project a reality.

Jorge Faytong performed some dataset collection/polishing, backend programming, and wrote this report.

Tomas Lampo did the user’s manual, some dataset polishing, and most of the backend programming.

Kotaro Hara and Michael Whidby were in charge of the interface development, visualizations, and conducting the tool’s usability test.

8. REFERENCES
Figure 6: Application Screen shot