Understanding and Supporting search for scholarly knowledge

Marcos Baez, Daniil Mirylenka and Cristhian Parra
The Problem
The Goal

To understand and improve the way find Scientific Knowledge

Understanding

Supporting
Understanding

Qualitative

30 Semi-Structured Interviews
Analyzing references on authored papers

Quantitative

Online Survey based on qualitative results
Qualitative Study

30 Semi-Structured Interviews

For each citation of one authored paper:

How did you get to this paper?

“My advisor suggested it”
“I search for X on Google Scholar”
“I was following some citations in a paper”
~ 700 interview notes. ~ 250 references. ~16% of Self citations.
Qualitative Results

Social Networks

- community/field: 18.06%
- colleagues: 15.48%
- venues: 14.19%
- collaboration: 12.90%
- senior colleagues: 10.97%
- coauthors: 10.32%
- research group: 7.10%
- acquaintances: 5.81%
- friend: 3.23%
- educational: 1.94%
Quantitative Study

How do you find interesting papers?

As part of our research at the ICT School of the University of Trento, we are conducting an experiment to analyze and understand how researchers stumble on scientific literature that they later end up citing in their work. Only aggregate results will be released once available, from this web site.

Let's browse through some of your papers. Please enter your name: ____________ Let's Go!

Interested on our work? Please, write an email to us

http://survey.mateine.org

~ 5 x 10^6 publications with references from Microsoft Academic Search
Quantitative Results

~ 226 References. ~ 23 different publications. ~ 25 different authors
Quantitative Results

Social Networks
Quantitative Results

- Search: 34%
- Social: 49%
- Navigation: 11%
- Other: 6%

Liked References
Our Understanding

Finding scientific knowledge has a strong social component to be leveraged
Supporting: Our Goal

Incorporating the social component in scholarly search by using the knowledge of researcher’s social networks

Social-Aware Search Results
Experiment setup

Sample of 1000 researchers

T = year 2005

“past” “future”

Researcher’s network at time T

“Past” references of the network

Top k recommendations

1.
2.
3.
...
k.

Researcher’s “future” references

Calculate:

precision
recall

depending on
the number of recommendations
popularity measure
Experiment setup

Popularity metrics:
- **network** – papers most cited in the network,
- **overall** – papers most cited overall,
- **random** – random set of papers

Performance measures

<table>
<thead>
<tr>
<th></th>
<th>Precision</th>
<th>Recall</th>
</tr>
</thead>
<tbody>
<tr>
<td>“Past” references of the network</td>
<td>(\frac{</td>
<td>C</td>
</tr>
<tr>
<td>Recommended papers</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

“Future” references of the researcher
Results

11% = P(you will cite the paper most cited between your coauthors)
20% of the papers you are going to cite have been cited by your coauthors

Ideas:
- Limiting the number of coauthors
- Weighting the recommendation with the number of coauthored papers

DID NOT improve the precision
Thanks for your Attention