Evaluation of Contextualization and Diversification Approaches in Aggregated Search

Hermann Ziak, Roman Kern
TIR 2017
1. Introduction
2. System
3. Evaluation
INTRODUCTION
Starting Point & Goals

- Provide & research ways to include *diversity* and *serendipity* into *result aggregation*
  - … specific for our search scenario
  - … for a better understanding of how users perceive such result aggregation techniques
Diversity

- Diversity could be described as a “strategy to resolve the ambiguity of a query”

- Usual approach:
  1. Capture the different meanings of a query
  2. Produce results that respond to each of the different meanings
INTRODUCTION

Serendipity

• “Serendipitous information retrieval is the occurrence of a user interacting with an information node with no prior intentions to do so”

• Pleasantly surprising results
  • Support a creative process

Cross Vertical Aggregated Search

- Federated Search consisting of heterogeneous sources
  - E.g. text, images, videos, …
Cross Vertical Aggregated Search

- Main challenges
  - Collection representation
  - Collection selection
  - Result merging
    - How to aggregate results from multiple sources & queries


INTRODUCTION

Scenario

- Cultural Heritage

Information Need

- Users are assumed to be in a state of a creative process
- Users have different background & knowledge

Query Formulation

- Queries are automatically generated by the user context
- “Just-in-time Information Retrieval”

SYSTEM
EEXCESS System

- Technological base for our adaptations
- **Federated recommender system**
  - ... developed for the cultural heritage domain
- Provides components for
  - E.g. query generation, federated search, result aggregation, de-duplication...
- Provides many search engines
  - E.g. Europeana, Mendeley, EconBiz, ...
- Open source
  - [https://github.com/EEXCESS](https://github.com/EEXCESS)
SYSTEM

Example Screenshot of the EEXCESS System
Approach

1. **Extend** the EEXCESS system to **use different methods** of including diversity and serendipity into the result aggregation

2. **Measure** the how well the different results aggregation methods are perceived
Three main approaches for result aggregation

- Non blended
  - Individual result lists

- Blended
  - One integrated result list

- Composite retrieval
  - Mixture of both


Diversity

- Query expansion
- Pseudo-relevance feedback via external source
- Disjunction query

Serendipity

- Query expansion
- Keyword extraction from user browsing history
- Conjunction query


# SYSTEM

## Comparison of Search Results from Wikipedia

### Search Query 1: einstein OR (albert OR bose OR relativity OR nuclear)

**Albert Einstein**

Albert Einstein (14 March 1879 – 18 April 1955) was a German-born theoretical physicist.

**Einstein** developed the theory of relativity, one of the two pillars

- **141 KB (15,113 words)** - 20:43, 25 August 2017

**List of scientific publications by Albert Einstein**

Albert Einstein (1879–1955) was a renowned theoretical physicist of the 20th century, best known for his theories of special relativity and general relativity

- **161 KB (4,081 words)** - 18:56, 31 July 2017

**Albert Einstein Science Park**

the park are the Einstein Tower, an astrophysical observatory that was built to perform checks of Einstein's theory of General Relativity, and the Great

- **2 KB (206 words)** - 20:59, 30 November 2016

**Albert Einstein House**

The Albert Einstein House at 112 Mercer Street in Princeton, Mercer County, New Jersey, United States was the home of Albert Einstein from 1935 until

- **6 KB (502 words)** - 00:26, 30 June 2017

**Satyendra Nath Bose**

Saha, Bose prepared the first book in English based on German and French translations of original papers on Einstein's special and general relativity in

- **28 KB (2,810 words)** - 18:14, 16 August 2017

### Search Query 2: einstein AND (music OR sports)

**Albert Einstein**

Albert Einstein (14 March 1879 – 18 April 1955) was a German-born theoretical physicist.

**Einstein** developed the theory of relativity, one of the two pillars

- **141 KB (15,113 words)** - 20:43, 25 August 2017

**Arik Einstein**

pionner of Israeli rock music and was named "the voice of Israel". Through both high public and critical acclamation, Einstein is regarded as the greatest

- **24 KB (2,931 words)** - 19:32, 2 August 2017

**Satyendra Nath Bose** (section Bose–Einstein statistics)

early 1920s, providing the foundation for Bose–Einstein statistics and the theory of the Bose–Einstein condensate. A Fellow of the Royal Society, he was

- **28 KB (2,810 words)** - 18:14, 16 August 2017

**Baby Einstein**

they named Baby Einstein and later sold as Language Nursery. The original video shows a variety of toys and visuals interspersed with music, stories, numbers

- **19 KB (2,120 words)** - 14:00, 24 August 2017

**Albert Einstein in popular culture**

Albert Einstein has been the subject of or inspiration for many works of popular culture. On Einstein's 72nd birthday on March 14, 1951, United Press photographer

- **23 KB (2,349 words)** - 23:41, 23 August 2017
Three types of results

- Results from the original query

- Results from the diversity query

- Results from the serendipity query

Note: all result lists are filtered for (near) duplicates
5 types of result aggregation

- Baseline
- Interleaved
- Three blocks
- Two blocks (diversity)
- Two blocks (serendipity)
5 types of result aggregation

- Baseline
- **Interleaved**
- Three blocks
- Two blocks (diversity)
- Two blocks (serendipity)
5 types of result aggregation

- Baseline
- Interleaved
- Three blocks
- Two blocks (diversity)
- Two blocks (serendipity)
5 types of result aggregation

- Baseline
- Interleaved
- Three blocks
- **Two blocks (diversity)**
- Two blocks (serendipity)
5 types of result aggregation

- Baseline
- Interleaved
- Three blocks
- Two blocks (diversity)
- Two blocks (serendipity)
EVALUATION
How to measure?

• “Cranfield paradigm” does not appear to be applicable
  • E.g. MAP, NDCG
  • … only focus on relevancy

• Measures to estimate the diversity exist
  • E.g. NDCG-IA, α-NDCG

• …still, not clear how to create the ground truth


Crowd Sourcing

- **Platform:** CrowdFlower

- Human workers were given tasks to assess
  - Over 300 workers
  - Over 1500 judgments

- Queries generated out of a query log
  - Total of 52 queries (from the cultural heritage domain)


Tasks consisted of comparing two result lists

- Vote for the preferred result list
- Background given for each query
  - Plus a Web search link
- Each task conducted at least by 6 workers
- Left/right swapped
- Static result lists (canned results)
- Methods to detected “cheaters”
EVALUATION

Evaluation scenarios

• Scenario #1
  • Baseline vs. interleaved | blocked

• Scenario #2
  • Baseline vs. two blocks (diversity | serendipity)

• Scenario #3
  • Interleaved vs. blocked

Measures

• Agreement on task level
• Ratio of “wins”
## EVALUATION

<table>
<thead>
<tr>
<th>Scenario #1</th>
<th>Agreement</th>
<th>Wins</th>
</tr>
</thead>
<tbody>
<tr>
<td>Interleaved</td>
<td>69%</td>
<td>64%</td>
</tr>
<tr>
<td>Blocked</td>
<td>72%</td>
<td>64%</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Scenario #2</th>
<th>Agreement</th>
<th>Wins</th>
</tr>
</thead>
<tbody>
<tr>
<td>2 blocks (diversity)</td>
<td>77%</td>
<td>69%</td>
</tr>
<tr>
<td>2 blocks (serendipity)</td>
<td>75%</td>
<td>69%</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Scenario #3</th>
<th>Agreement</th>
<th>Wins</th>
</tr>
</thead>
<tbody>
<tr>
<td>Blocked vs. interleaved</td>
<td>65%</td>
<td>47%</td>
</tr>
</tbody>
</table>
How confident were the workers?

<table>
<thead>
<tr>
<th></th>
<th>Interleaved</th>
<th>Blocked</th>
<th>2 blocks (diversity)</th>
<th>2 blocks (serendipity)</th>
<th>Blocked vs. interleaved</th>
</tr>
</thead>
<tbody>
<tr>
<td>Very Confident</td>
<td>30</td>
<td>34</td>
<td>42.5</td>
<td>47.5</td>
<td>27</td>
</tr>
<tr>
<td>Confident</td>
<td>52</td>
<td>47</td>
<td>41.5</td>
<td>38</td>
<td>59</td>
</tr>
<tr>
<td>Hard</td>
<td>18</td>
<td>19</td>
<td>16</td>
<td>14.5</td>
<td>18</td>
</tr>
</tbody>
</table>
**Insights**

- Each configuration better than the baseline
- Diversity and serendipity similar
  - Depends on the query, which is preferred
- Two blocks preferred over three blocks
- Need to take the background of the user into account
  - Web search were often used
CONCLUSIONS

- Either use diversity or serendipity
  - ... depending on the query
  - ... and (potentially) the user

- More research on serendipity is needed
  - ... especially improved evaluation scenarios

- Results & data set available on Github
  - https://git.io/vQTLT