Challenges in disseminating Scientific and Cultural Resources in the Long Tail

Prof. Dr. Michael Granitzer
University of Passau
26.03., Hamburg, Germany
Agenda

• Our Content Distribution Process today
  – The Long Tail of Information
  – Popular vs. Scholarly
  – The Filter Bubble – Rich get Richer – Popularity ranking

• EEXCESS – personalised just-in-time recommendation of long tail content
  – Federated Recommendation
  – User and Usage Mining
  – Discovery via visualisations

• Summary
Motivation
Availability of Digital Resources

• Vast amounts of scientific resources digital available

• Content distribution processes optimized towards highly popular, commercial media content
  – Niches do play less of a role
  – Scientific content **not main stream**
  – Remains hidden in the **Deep Web**
  – Hard to **discover** unknown items

• Lets have a look in detail, why that is the case
Challenge
The long tail of websites

Reach of Sites on the Internet

- Search Engine Optimization
- Social Media Marketing etc.
Challenge
The content distribution process today

- Discovering non-main stream cultural, scientific or educational content is challenging
- Expertise of memory organisation remains untapped, since not part of the distribution process
- Users are disconnected from the experts
Challenge
Scholarly vs. Popular Resources

http://southwestern.edu/su_blogs/library/files/2012/02/scholar-v-pop-flierTZ.jpg
**Challenge**
What is wrong with popularity?

**Phenomena**
- Rich-get-Richer effects
  - Hard to become rich when starting in the long tail
- Potential Filter Bubble
- Popularity can make rankings unfair
Challenge
Is popularity based ranking the best we can do?

Experiment by Salgankik, Dodds, and Watts

• Download site for 48 obscure music songs
  – User get feedback on the popularity of a song
  – Hidden to user: 8 different copies of the system
    ➔ 8 very different market shares
  – 9th System: no user feedback
    ➔ Less variation in market shares of different songs

➔ Popularity reinforce rich-get-richer phenomena
DISCOVERING SCHOLARLY & CULTURAL CONTENT IN THE LONG TAIL
Project Details

- EU FP7 funded Integrated Project
- 10 Partners
  - R&D Partners
  - Memory Organisations

Vision: Bring the user to the content, not the content to the user
A Quick Introduction
Idea

„Bring the content to the user, not the user to the content“

• Inject cultural and scientific content into existing web channels
  – Websites (Wikipedia, etc.)
  – CMS/LMS
  – Social media channels (Twitter, etc.)
  – Support “head-channels” as well as tail-channels

• Gather user and usage feedback such that memory organisations can learn how resources are distributed and used

Requires high-quality, context dependent injections of resources (aka personalised recommendations)
Approach Overview

Content Consumption (e.g. Browsing, SNA)

Content Creation (e.g. Writing Blogs, Editors)

Involving in Content Consumption

Involving in Content Creation

Recommendation

Content

ZBW Content
Mendeley Content
AMBL Content
Europeana
CT Content
Open Access

Context

content

content

content
Approach

Example Wikipedia
Approach
Example Blog Writing

Last night I have discovered Jupiter standing unusually high in the night sky. The explanation for this phenomenon: Jupiter has reached opposition (directly opposite the Sun, as seen from our beloved Earth) the other day. This year’s opposition is particularly favourable for us here on the Northern hemisphere, with Jupiter climbing high in the Southern sky at midnight.

Even the four largest moons of Jupiter, called Galilean Moons (discovered by Galileo Galilei in 1610) were to be seen. The derive their names from the lovers of Zeus in the Greek mythology: Io, Europa, Ganymede and Callisto.

Available Recommended Links:
- Galilean moons
- Jupiter
- Earth
- Sun
- Night sky
Approach
Federated Recommendation

1. Select appropriate source
2. Forward request
3. Gather and merge results
4. Return results
Approach
User and Usage Mining

• Obtaining knowledge over user and usage while retaining privacy
Approach
Intelligent User Interfaces

- Beyond 10 results through visualisation
  - Recommendation Results
  - Relationships among items
  - Metadata
- Empower user to discovery interesting resources in the long tail content
Approach
Intelligent User Interfaces

- Facetted browsing of large result lists for discovery purpose
Approach
Intelligent User Interfaces

• Query Summary
Approach

Intelligent User Interfaces

• Explorative visualizations to foster discovery
  – Temporal vs. Topics
  – Citation Graphs
Approach
Evaluation in 3 Testbeds

Educational Support
- Cultural/scientific resources injected to LMS
- Pupils, Teachers

Scholarly Communication
- Interconnecting cultural and scientific resource
- Students, Lecturers, Researchers

General Public Education
• Disseminate cultural/scientific content to the general public
• Regionally interested users, culturally interested users, media consumers
Outcome
New content distribution processes

- **Empower** memory organisations to shape their content distribution process and bring in their expertise
- **Enrich** existing content consumption and creation processes (e.g. blogging)
- **Personalized** recommendation of relevant content based on user and usage analysis
- **Transparent and trusted** distribution processes
Future Vision: Scale beyond EEXCESS

- **Adoption** by a large number of different memory organisations
- **Enriching** the Web with cultural, educational and scientific content
Summary

Popular vs. Scholarly
- Long tail content matters
- Challenge in overcoming the content distribution process today

Bring the content to the user
- Federated Recommendation
- User and usage mining while preserving privacy
- Intelligent user interfaces to foster discovery

Connecting users and experts (again)
In theory there is no difference between theory and practice. In practice there is. (Yogi Berra)