CROWDSOURCING TO MAKE PICTURES VISIBLE FOR RESEARCH

Majlis Bremer-Laamanen

LIBER 5TH OF OCTOBER, 2011
WHY? GOALS

• Libraries are digitising huge corpuses of collections.

• Users have an increasing demand for more sophisticated and diverse search functionalities.
  • Needs of researchers
  • Funding situation

  – Users want to be involved with the content:
    • To contribute to it. To do something important. To be a part of a larger goal
    • Rewards by the content, game, scoring…
TO THE FUTURE
FROM THE PAST
Customer
Mobile
PC
Kindle
Sony….

Advertisements
Articles
Photos, pictures
Captions
Cartoons…

Development issues

World Digital Library

National Digital Library

Digi.nationallibrary.fi

Collections
- Open Access
- Creative Commons
- Licensing

NDL
Europeana

Other digital libraries

Europeana

Google

Social media

Crowdsourcing

Research and collaboration, e.g.
Helsinki University (language and IT),
Aalto University (ontologies)
and
Tampere University (ocr)
ENRICHING CONTENT

Statistics

• 2010: 6.5 million page views
• Newspapers 1771-1910: 2 million pages
• Journals: 1810-1910 -1944 2.5 million pages
• Industrial prize lists (1810-1944: 129 000 pages)
• Books, audio: 2011

Interest groups

• The creators of the original materials
• The users of the digitised materials: e.g. researchers, students, the general public – user from all around the world
• The contributors: the reuse and enriching of the digitised materials

Digi.nationallibrary.fi

The national identity on the Web

Vast numbers of newspapers, periodicals, small printed items and books contained in the rich collections of the National Library of Finland have now been digitised to form a fascinating material that throws light on the past and presents as parts of our national identity.

The opportunity to browse through and search this material is now within the reach of scholars, students and all users of the Web.

Newspapers published today 150 years ago 10.05.1860

• Abo Underrättelser
• Finlands Allmänna Tidning
• Suomen Julkisia Sanoma

http://digi.nationallibrary.fi

The earlier version of this service is still available here.
KUVATALKOOT

KUVA = IMAGE / PICTURE / PHOTO / ARTICLE
TALKOOT = PEOPLE GATHERING TO WORK TOGETHER VOLUNTARILY (WITHOUT PAYMENT)

FIRST EXPERIENCE 2011:
DIGITALKOOT: correction of OCR by gamification, turning useful activities into games ”THE MOLE HUNT” by Microtask.com.

From February, 8th to September 15th, 2011: about 80,000 visitors, 4000 hours of effective game time. More than 5 millions of tasks.
DIGITALKOOT

• In the Helsingin Sanomat news 15.9
• Digitalkoot and HS Talkoot
• Helps the library to interpret the OCR:d text in fraktur
• The participants:
  – Help the moles to build a bridge
  – To plant flowers
• Result: more than 99% correct. The same word is checked by many participants
Published cultural heritage:
Funding: European Social Fund 2011-2013
160,000 euros.
Goal:
- Extend crowdsourcing to images and/or articles
- To use keywords and/or article headings
- Focus important
Aim of the project:

1) Marking of articles and/or images

2) Indexing articles and/or images
Goals: sophisticated user experiences

Collections discovery and reuse of digital content by researchers and people at large:

Researchers will get better systematic coverage of images and articles in published printed material.
The platform

- Usability and seamless access
- Stand-alone or integrated model?
- Game or other possibilities?
- Rewards for the user
USER COMMUNITIES/VOLUNTEERS

- At first: Regional testing with students, researchers of history…
  - The knowledge competence influences the need for quality and the aim of the work
- Facebook and internet use
- The results will be taken into account when planning the user portal to be.
Enriching Digitisation Production Processes, METS Profiles: a new development platform

**SOURCE MATERIAL**

**PHYSICAL COLLECTIONS**

- Structural metadata
- Administrative/technical metadata
- Descriptive metadata

**CATALOGUING**

- Articles
- Illustrations
- Poems

**SCANNING**

- METS
- ALTO

**POST PROCESSING**

- METS, ALTO
- MIX/PREMIS
- MARC21/MODS

**LEVEL OF MARK UP**

- Articles
- Illustrations
- Poems

**DIGITAL RESOURCE**

**COMPREHENSIVE DIGITAL COLLECTIONS**

- Standards & OAI-PMH compliant METS SIP packages

**METS EXPORT Packages include:**

- JPEG2000
- OCR TXT as ALTO XML
- PDF
- J PEG(150)
- METSXML
- MARCXML

**LEVEL OF MARK UP**

- Newspapers
- Serials
- Books
- Parchments
- Notes
- Maps
- Audio

**Two Bibliographic Records**

- Articles
- Illustrations
- Poems
In the Kuvatalkoot we are engaging curious and enthusiastic volunteers.
Planning part II

…without forgetting the valuable nature of the content.

Voima 4/1922