

# Two Years on - Developing Metrics for Crowdsourcing with Digital Collections

Tuula Pääkkönen, Information Systems Specialist,  
6th International Conference on Integrated Information,  
Athens (virtual presentation)

21.9.2016

# National Library of Finland



Centre for  
Preservation &  
Digitisation

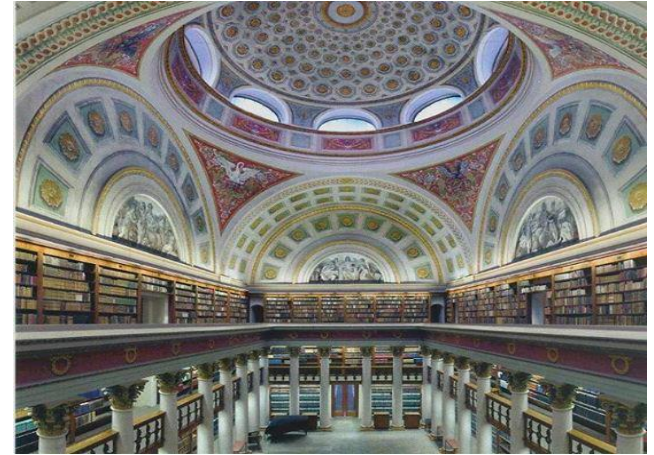
**2015**

**2006**

**1990**

**1828**

Imperial Alexander University



**1707**

Free deposit law

**1640**

Turku Academy



Akatemiatalo. Akademiäuset. The Academy. C.C. Gjörrwell. 1815.



THE NATIONAL LIBRARY OF FINLAND

Source: <http://www.kansalliskirjasto.fi/yleistieto/kirjastotietoutta/historia.html>



# National Library of Finland and Crowdsourcing



# The Task for Crowdsourcing - Find articles

1640 DIGI.KANSALLISKIRJASTO.FI FESTIVALYLEIN

FRONTPAGE NEWSPAPERS JOURNALS EPHEMERA OTHER DIGITAL COLLECTIONS SCRAPBOOK

SUOMENKI  
PÄ SVENSKA  
IN ENGLISH

Tuula

Feedback

MAURY & CO. HUISTIKKI.  
Pääsiamehiet.

**Uleiset Pääsiäis- Walwojaiset**  
Paloikunnan talolla 30 p. t. t.  
alkaa klo 8 ip.  
Waihteleva ohjelma!  
Esaraha 50 p. Kaikki terveh-  
tulleita!  
Työvä. Yhd. hupitoimikunta.

**Alatornion Nuorisoseura**  
toimii  
**Suvi-Yltaman**  
Huoneustossaan Raumolla Huhtik. 1  
p:nä. Ohjelmassa m. m. Näytöskappa-

**Alatornion Työvä-yhd.**  
toimii  
**Rufetti-Yltaman**  
Tän Pääsiäispäivänä klo 6 p. pp.  
Ohjelma: Laulu, Runo, Runonäytös.  
Huone: Keskisen katuja myöten lään-  
päätä.

**Tilauksia jalkineen valmistamiseen**  
vastaanotetaan Melubergin talos-  
sa Torniossa.  
H. M. Penttinen.

**Leimonen.** Viime sunnuntaina  
kuin jalkineet hiihtelijät näkimät  
he suuresti ihmettelevän Leimonen n. t.  
Kierimäoralla liivertelevän. Todek-  
latin ihmeellistä tähän vuoden ai-  
kaan.

**Toimikunta.**  
\*\*\*\*\*  
**Salapoliisi Holmes**  
salaisuus  
eli  
**Tolsen Tahrans Tarina.**  
Kirj. Conan Doyle.  
Jatk.

John Mittonin vangitsemisen  
tehtiin vain epätoivossa siitä, et-  
tei voitu mitään muutakaan teh-  
dä. Häntä ei voitu syyttää mis-  
tään. Hän oli ollut sinä iltana  
erään ystävänsä luona Hammers-  
mithissä; hän oli täysin syytön  
Hän oli tosin lähtenyt kotimat-  
kalle ennenkuin rikos huomattiin,  
mutta hänen oma selityksensä,  
että hän osan matkasta oli kul-  
kenut jalkaain, pidettiin todenmu-  
kaisena, kun ilma sinä yönä oli  
kannis. Hän oli saapunut kello  
12 ja näkyi joutuneen aivan hä-  
milleen huomattavissa tapahtu-

lassa rue Ansterl...  
Pariisin viranomai...  
oli tullut mielenvil...  
osottikin, että hän oli  
nyt järkensä ja joutui  
johteen valtaan. Poliis...  
nut selville, että rouv...  
Fournaye aivan asketta...  
punnut Lontoosta, jossa...  
nantaina oli ollut käym...  
nen matkansa on todia...  
yhteydessä Westminsteria...  
tuneen rikoksen kanssa...  
vien vertailu on lopullis...  
tanut, että hra Henry...  
ja Eduardo Lucas on ol...  
henkilö ja jostain syyt...  
nyt kaksoiselämää Lon...  
Pariisissa. Rouva Fourn...  
on kreolilainen syntään...  
teeltään äärettömän tur...  
ja viime aikoina on ha...  
vannut mustasukkaisuus...  
li muuttanut suoranais...  
voksi. Arvellaan hänen...  
sa tilassa tehneen murh...  
nykyään valtaa mielet L...  
Missä hän on liikkunut...  
vasten yöllä, sitä ei ole...  
tu saamaan selville, mut...  
tenkin tiedetään, että rouva For-  
naye'n näköinen nainen tiistai-  
muna herätti suurta huomiota

Create a clipping

Title \*  
Salapoliisi Holmesin salaisuus eli Tolsen Tahrans

Subject  
Art. culture and entertainment

Category \*  
Novel, short story, story, description

Keywords  
Arthur Conan Doyle Sherlock Holmes

Cancel Save

# Users for Crowdsourcing

Mostly same as the "browsing" users

- Family researchers
- Local historians
- Academics
- Teachers, school classes
- Hobbyists of various interests (cottage renovators)

# Why Crowdsourcing?

- To get to know the users and their interests
  - Which content is the most interesting
- Improving quality of the collection
- Improving quality of the metadata

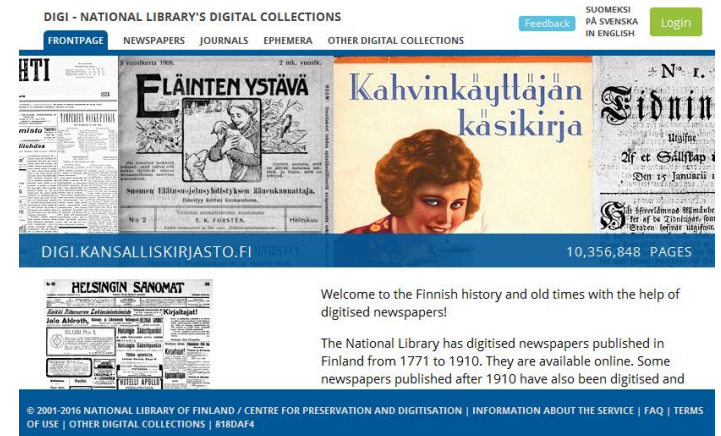
"Hi, I'm searching information of ... How do I ..."

"Hi, there is a page missing"

"Hi, on that year this newspaper was not published on this day."

# Beginning of Crowdsourcing

- New version was launched 13.5.2014 which had the crowdsourcing functionalities available

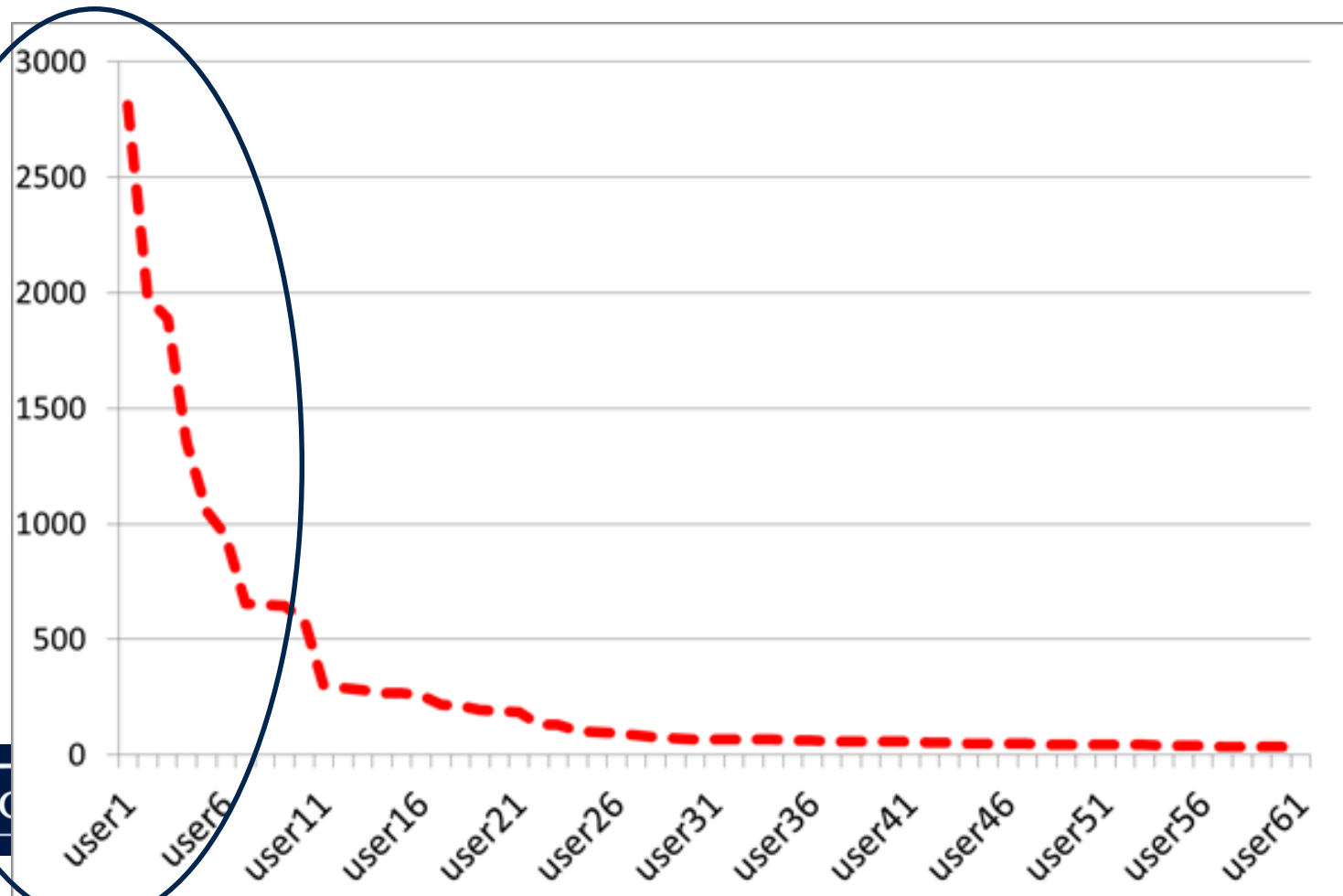




# Crowdsourcing metrics

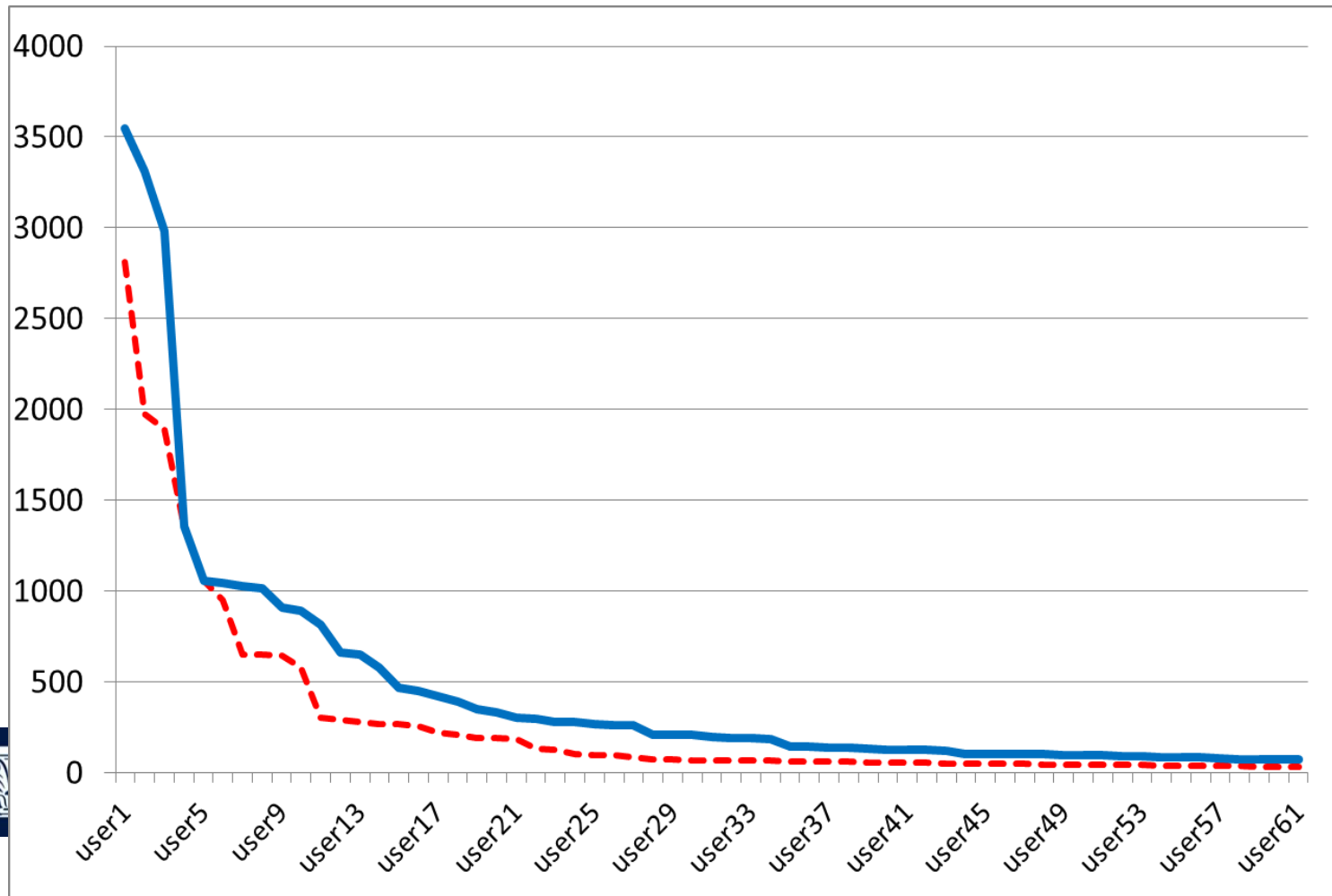
# Clippings Quantities on first year

- Clippings were created gradually
- Already after first months there were few very high top users.



# After 2nd year (this year June)

- Blue line is the situation after 2nd year of crowdsourcing
- Red line is the situation after 1st year of crowdsourcing



# Data of Clippings Quantities

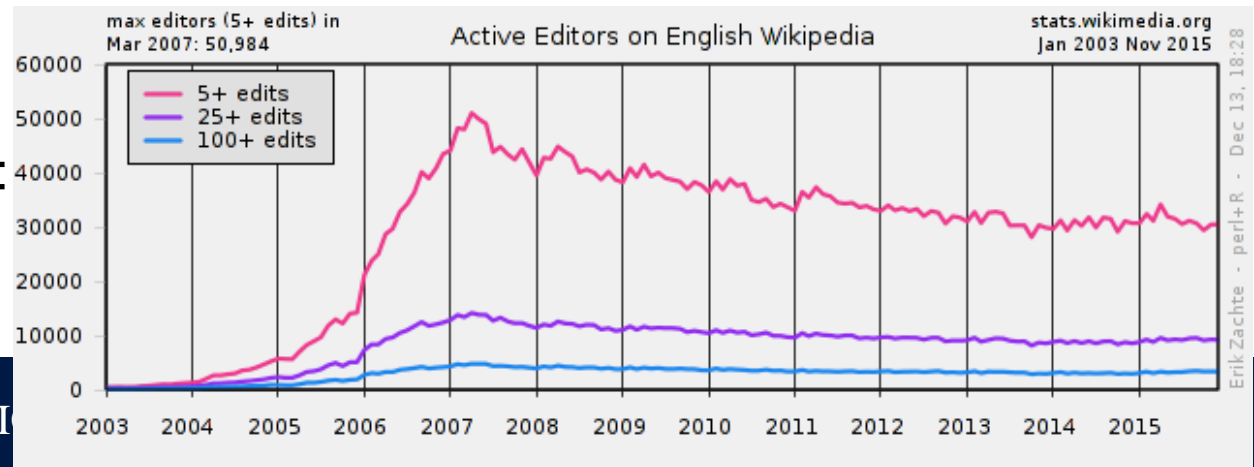
Clippings data	1st year	2nd year
Amount of clippings	19666	35345
Top-5 (%)	46 %	35 %
Top-10 (%)	64 %	49 %
Mean	44,49	40,91
Maximum	2811	3550

# Trends

- It is very beneficial to have some "hyper-active" users crowdsourcing-wise, as they can cover much of the work done.
- Same kind of situation as in wikipedia and in the case of the Australian Digitized Collections, Trove  
<http://trove.nla.gov.au/system/stats?env=prod#corrections>

Image source :

[Wikipedia](#)





# Crowdsourcing metrics and themes of clippings

# Most used keywords in clippings

Keyword	English translation	1 <sup>st</sup> year (2015)	2 <sup>nd</sup> year (2016)	Difference
<b>tanssi</b>	dance	1686	3313	1627
<b>automobil</b>	car (old format, in Swedish)	1089	1166	77
<b>automobiili</b>	car (old format, in Finnish)	998	1265	267
<b>tanssinopetus</b>	dance teaching	913	2492	1579
<b>tanssikurssi</b>	dance course	806	904	98
<b>tanssikoulu</b>	dance school	497	1304	807
<b>negatiivinen suhtautuminen tanssiin</b>	negative outlook on dancing	445	468	23
<b>positiivinen suhtautuminen tanssiin</b>	positive outlook on dancing	324	341	17
<b>älytehtävä</b>	brain teaser/quiz	294	294	0
<b>korsord</b>	crossword puzzle (in Swedish)	190	190	0

# Most increased themes for 2nd year

Keyword	English translation	1 <sup>st</sup> year (2015)	2 <sup>nd</sup> year (2016)	Difference
<b>tanssi</b>	dance	1686	3313	1627
<b>tanssinopetus</b>	dance teaching	913	2492	1579
<b>tanssikurssit</b>	dance courses	1	1480	1479
<b>tanssikoulu</b>	dance school	497	1304	807
<b>automobiili</b>	car (old format, in Finnish)	998	1265	267
<b>Kuolinilmoitus</b>	death notice	1	111	110
<b>tanssikurssi</b>	dance course	806	904	98
<b>Reo</b>	Reo (REO Motor company abbreviation)	53	138	85
<b>automobil</b>	car (old format, in Swedish)	1089	1166	77
<b>Sievin kauppiaat</b>	traders of Sievi	0	58	58
<b>Raamattu</b>	the Bible	2	55	53
<b>kolera</b>	cholera	125	175	50

# Limitations

- Via the free-to-use web system, the digitized material is available from years 1771-1910
  - most of the clippings also occur to that time period
- In recent project, Aviisi, we are working on opening in-copyright materials
  - Clippings have been surprisingly useful in education, for teachers and students
  - Also they are helpful for researchers

# Conclusions

- Finding and getting more of those hyper-active users, will benefit crowdsourcing
  - Where to get more users? (network, communities, organic growth)
  - Communications via projects etc.
- Metrics developed depend on goals set for the crowdsourcing





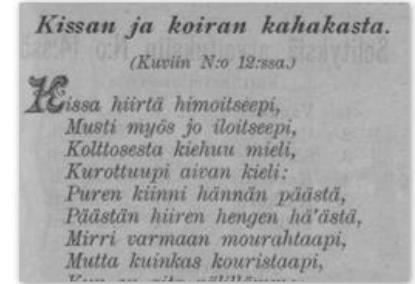
"Arkaillematta lähenin kis...  
Otavan joulukirjallisuus  
01.01.1904



Pieni hyväntekijä  
Kolito : Opettajayhdistyk...  
01.01.1901



Mirri ja mopet tuolla  
Kolito : Opettajayhdistyk...  
01.01.1896



Kissan ja koiran kahakasta...  
Kolito : Opettajayhdistyk...  
01.01.1896

# Thank you!

[Tuula.paakkonen@helsinki.fi](mailto:Tuula.paakkonen@helsinki.fi)

<http://digi.kansalliskirjasto.fi>

# More to read (see all from paper)

- [3] R. Holley, 'Crowdsourcing: How and Why Should Libraries Do It?', *D-Lib Mag.*, vol. 16, no. 3/4, Mar. 2010.
- [4] T. Pääkkönen, 'Crowdsourcing metrics of digital collections', *Liber Q.*, vol. 25, no. 2, p. 41, Dec. 2015.
- [5] J.-P. Hakkarainen, 'Introduction to Our Project | Fenno-Ugrica', *Fenno-Ugrica. The Blog of the Minority Languages Project – National Library of Finland*. .
- [6] T. Rose-Sandler, W. Ulate, M. Seidman, M. Flanagan, G. Belknap, V. van Hyning, and V. O'Donnell, 'Engaging the Citizen Scientist in Content Enhancement for BHL', presented at the Biodiversity Information Standards (TDWG 2015), Nairobi, 2015.
- [12] P. Karppinen, 'The Aviisi project: the golden century of newspapers put to new use | The National Library of Finland Bulletin 2016', *The National Library of Finland Bulletin* 2016, 03-Jun-2016. .
- [13] K. Kettunen, T. Honkela, K. Lindén, P. Kauppinen, T. Pääkkönen, and J. Kervinen, 'Analyzing and Improving the Quality of a Historical News Collection using Language Technology and Statistical Machine Learning Methods', presented at the IFLA World Library and Information Congress Proceedings 80th IFLA General Conference and Assembly, 2014.
- [14] P. Karppinen, M. Kaukonen, T. Pääkkönen, and M. Sorjonen, 'Contracts Enabling Collaboration of The National Library of Finland with Media Houses in Electronic Deposit', presented at the IFLA World Library and Information Congress, Columbus, Ohio, United States, 2016.