

User Search Terms and Controlled Subject Vocabularies in an Institutional Repository

Scott Hanrath | shanrath@ku.edu | @rshanrath

Erik Radio | radio@ku.edu

University of Kansas Libraries

What is this all about?

KU ScholarWorks and FAST

User Search Queries

User Queries → FAST

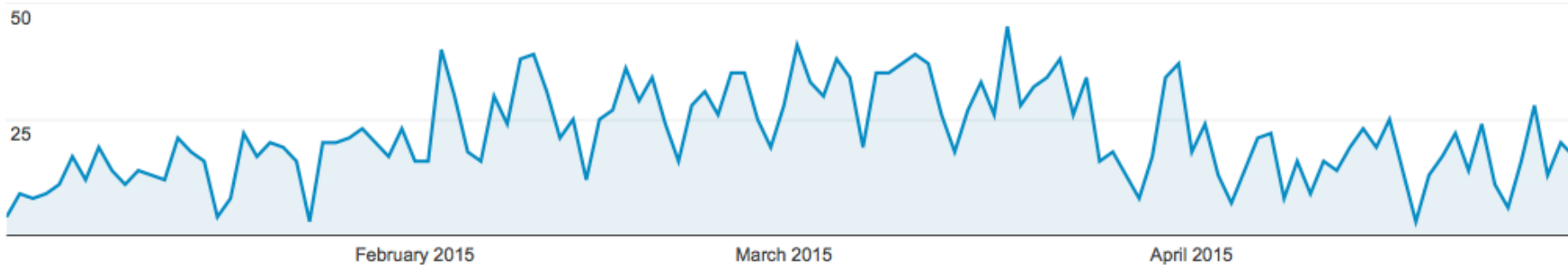
User Queries → Item Metadata

User Search Queries

Google Analytics Segment:

- Landing Page = Item simple record page
- Next interaction = “BitStream Click” Event
- Keyword from Search Source is set

● Total Events



| Page | Keyword | Total Events |
|---|---|---|
| Landed on handle with next interaction BitStream Click | | 2,609 % of Total: 15.49% (16,841) |
| 1. /handle/1808/11138 | peer tutoring | 22 (0.84%) |
| 2. /handle/1808/6891 | http://kuscholarworks.ku.edu/handle/1808/6891 | 21 (0.80%) |
| 3. /handle/1808/5341 | ravel personality | 10 (0.38%) |
| 4. /handle/1808/11138 | classwide peer tutoring | 8 (0.31%) |
| 5. /handle/1808/16245 | teaching older adults with learning disabilities | 8 (0.31%) |
| 6. /handle/1808/6223 | mental retardation | 8 (0.31%) |
| 7. /handle/1808/13065 | generic versus name brand | 7 (0.27%) |
| 8. /handle/1808/10945 | kids with disabilities going to school | 6 (0.23%) |
| 9. /handle/1808/5680 | http://kuscholarworks.ku.edu/handle/1808/5680 | 6 (0.23%) |
| 10. /handle/1808/6134 | clownfish and anemones | 6 (0.23%) |

User Search Queries

- 4 month period, January - April 2015:
- 2,209 Query – Item combinations
- ~ 15% of the total BitStream Click Events recorded

Reconciling Queries Against FAST

- Open Refine
- Ted Lawless' refine reconciliation service using the assignFast API:

<https://github.com/lawlesst/fast-reconcile>

<http://www.oclc.org/developer/develop/web-services/fast-api/assign-fast.en.html>

Reconciling Queries Against FAST

baba yaga



Baba Yaga (Legendary character)

INFORMATION ABOUT THE RESOURCE

SCHEMA.ORG NAME(S):

CONTROLLED HEADING IDENTIFIER:

<http://id.worldcat.org/fast/824930>

IDENTIFIER:

824930

TYPE:

[Intangible](#)

SKOS PREFERRED LABEL:

Baba Yaga (Legendary character)

SKOS ALTERNATIVE LABEL:

Baba Jaga (Legendary character)

IS IN SKOS SCHEME:

<http://id.worldcat.org/fast/ontology/1.0/#fast>

<http://id.worldcat.org/fast/ontology/1.0/#facet-Topical>



Image: http://commons.wikimedia.org/wiki/File:Bilibin._Baba_Yaga.jpg

of Queries Auto-Reconciled Against FAST

46 of 2209

(2%)

Manually-split Queries

- Sample of 300 queries.
- 97 (32%) were split.

cognitive disability and
internet

→ cognitive disability

→ internet

of Manually-split Queries Auto-Reconciled Against FAST

84 of 300

(28%)

Other Ways to Classify the Sampled Queries

- Known Item Searches
- Google Scholar Related Item Searches
- Query similarity to Item Metadata

Known Item Searches

52 of 300

(17%)

Google Scholar Related-Item Searches

(related:[some-id]:scholar.google.com/)

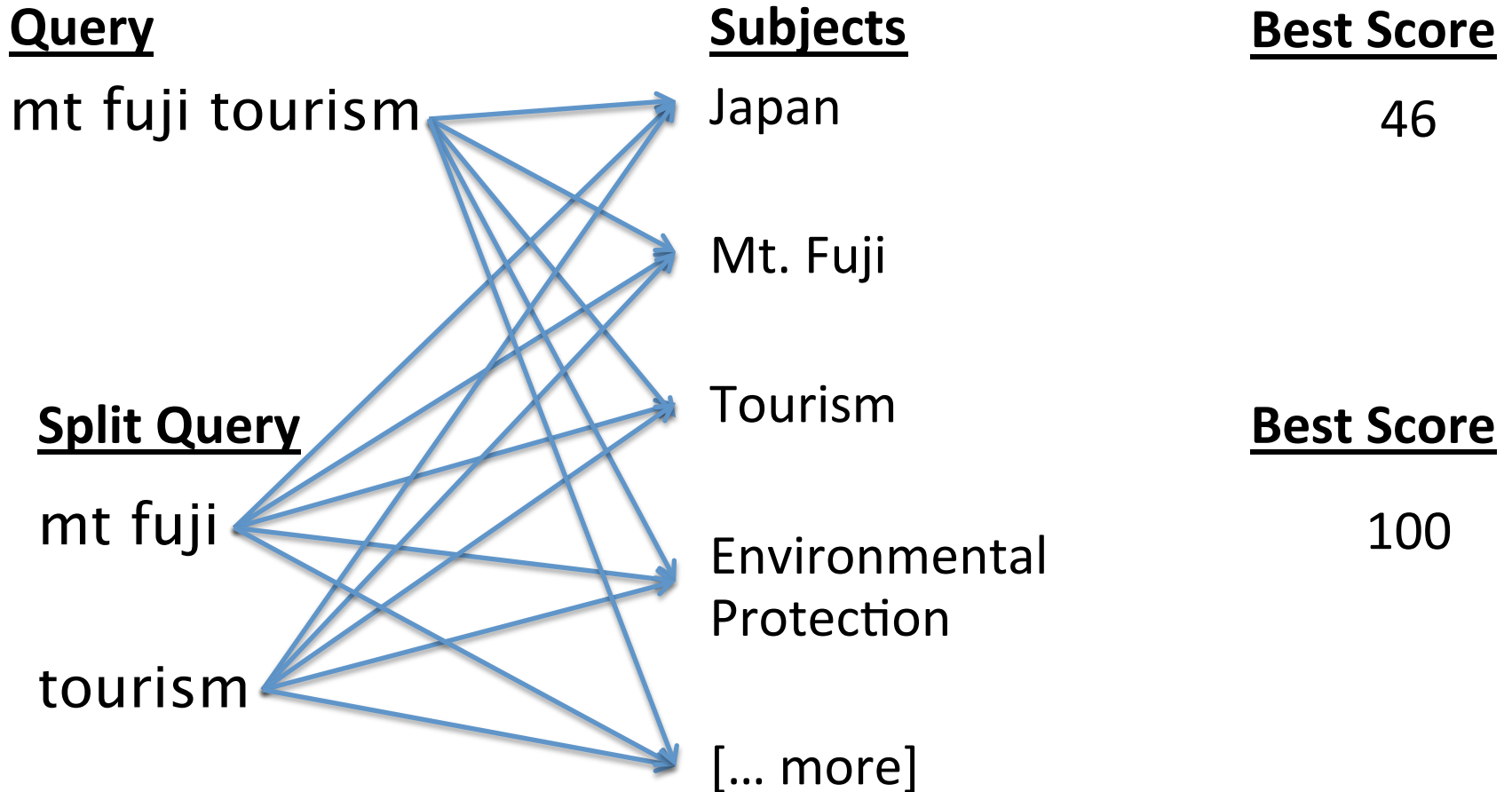
10 of 300

(3%)

Query Similarity to Item Metadata

- Fuzzy string matching: Levenshtein Distance between query and item metadata values for Title and Subjects
- Normalize score by string length, 0 to 100 scale (100 = same string)
- For multi-valued metadata fields, choose best match with query

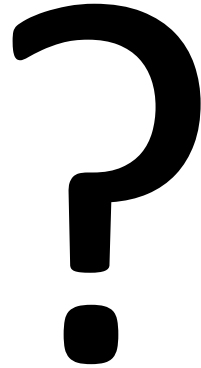
Query Similarity to Item Metadata



Query Similarity to Item Metadata

| Field | N | Not Very Good (< 75) | Pretty Good (75 to 95) | Very Good (95 to 100) |
|-------------------------------|----------|---|-----------------------------------|----------------------------------|
| Title | 298 | 91.6% | 3.0% | 5.4% |
| Subjects | 197 | 95% | 1.5% | 2.5% |
| Subjects, Split Queries | 197 | 87.8% | 3.0% | 9.1% |

Summary and Next Steps



Scott Hanrath | shanrath@ku.edu | @rshanrath

Erik Radio | radio@ku.edu

University of Kansas Libraries