



CC-SA-2.5

Knowledge Organization with Wikipedia – Joining the Free Encyclopaedia and Digital Libraries

Jakob Voss

Wikimedia Deutschland e.V. /
Common Library Network (GBV)

5th NKOS Workshop, Alicante
September 21, 2006

[[Outline]]

- * Wikipedia
- * Wikipedia's Category system
- * Mapping
- * Indexing with Wikipedia articles
- * Summary
- * Learn from Wikipedia

[[What Wikipedia is]]

- * Collaborative edited
- * Free Content
- * Encyclopaedias
- * ~5 million articles, >50.000 editors
- * 230 languages (~50 with > 10.000)
- * Non-commercial, no ads
- * TOP 16 website by access

[[How Wikipedia is created]]

- 1) Find something to improve
- 2) Do it right
- 3) Let other people improve it

[[Wikipedia categories]]

- * Tagging by linking

 - [[Categorie:Information Science]]

 - [[Category:Belgion Inventor]]

 - ...

- * Open for all

- * Blind tagging

- * No aggregation (set-model)

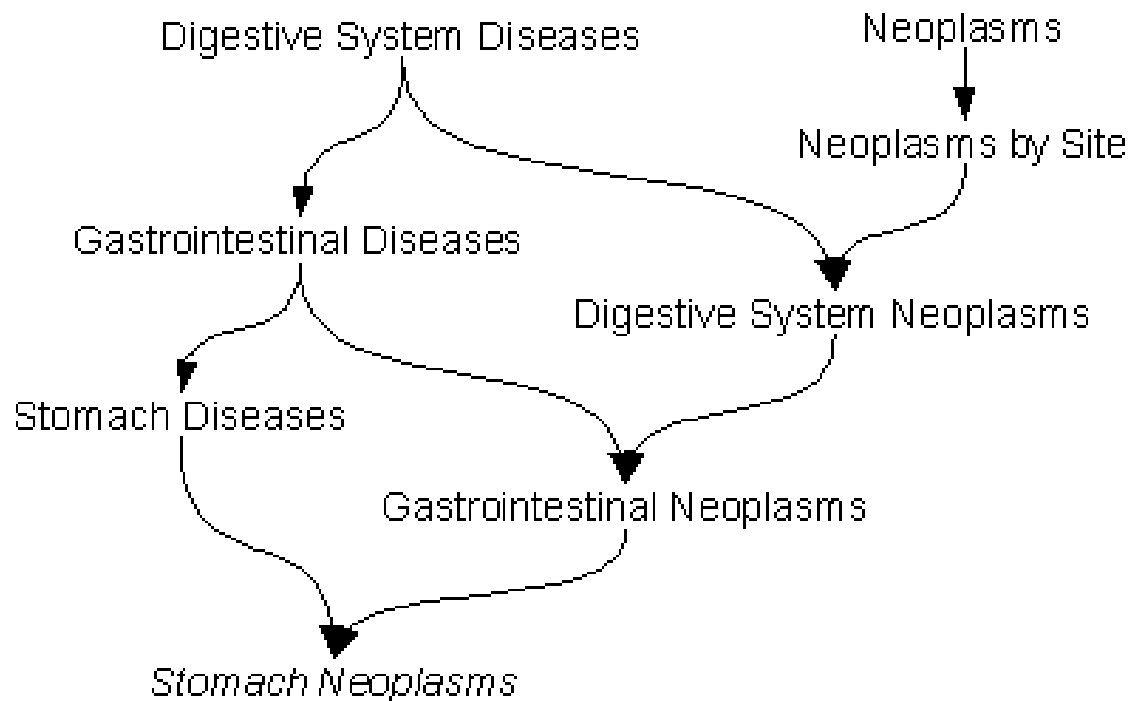
- * Multi-hierarchical relations

- * High connectivity

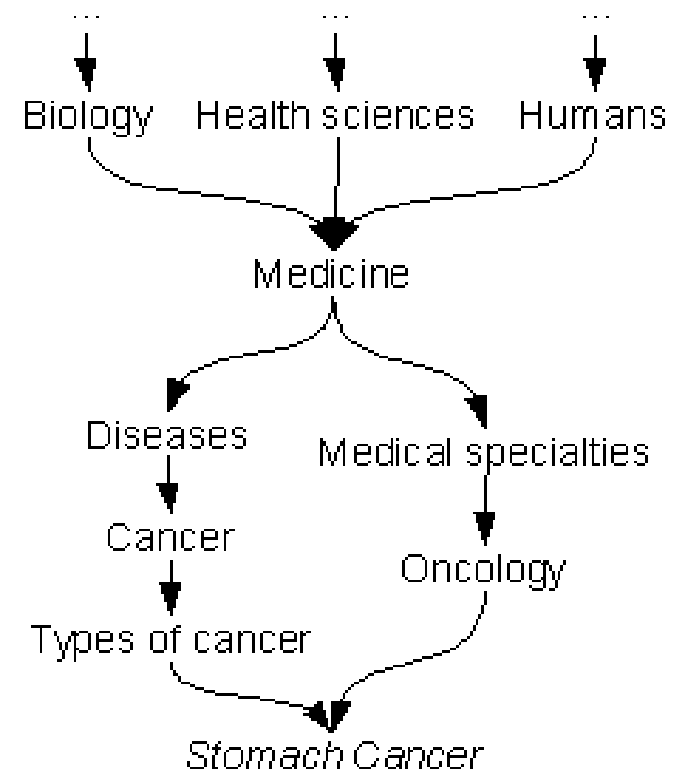
Example: http://en.wikipedia.org/wiki/Paul_Otlet

[[Wikipedia categories]]

Medical Subject Headings (MeSH)



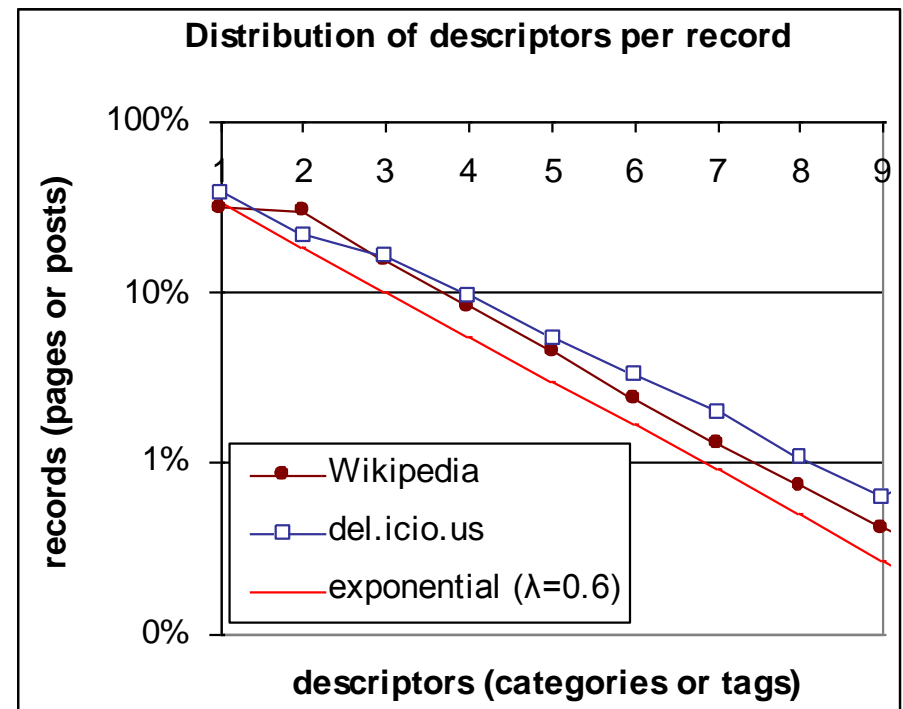
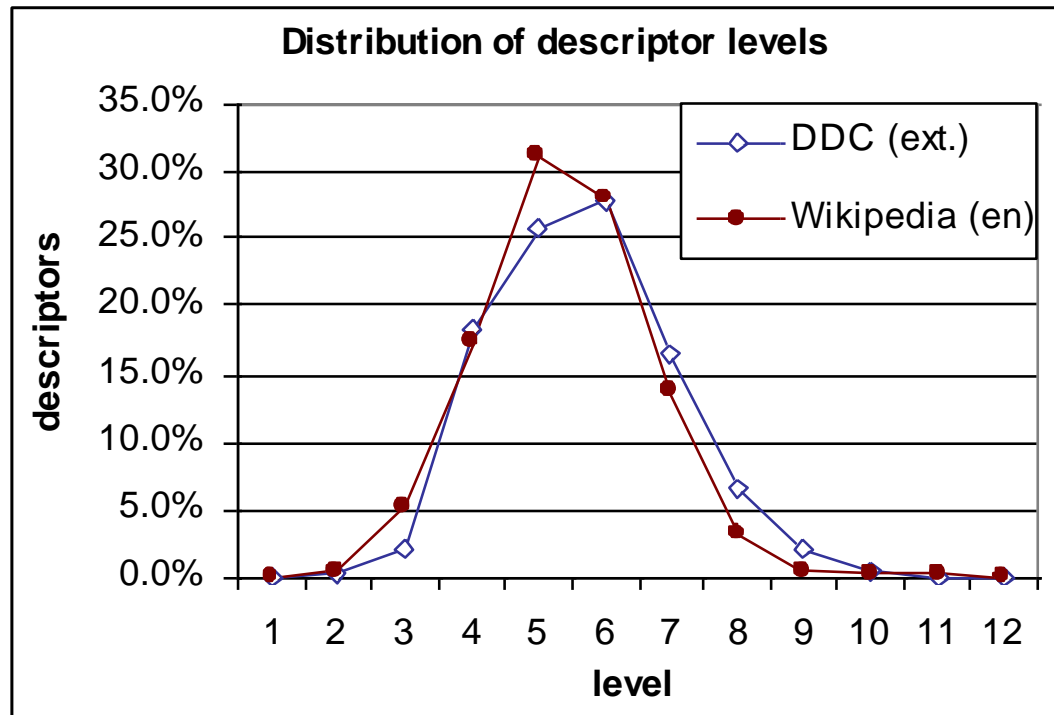
Wikipedia categories



[[Category system as KOS]]

Collaboratively edited, general thesaurus

(en, Jan 2006: 91,502 categories, 923,196 articles)

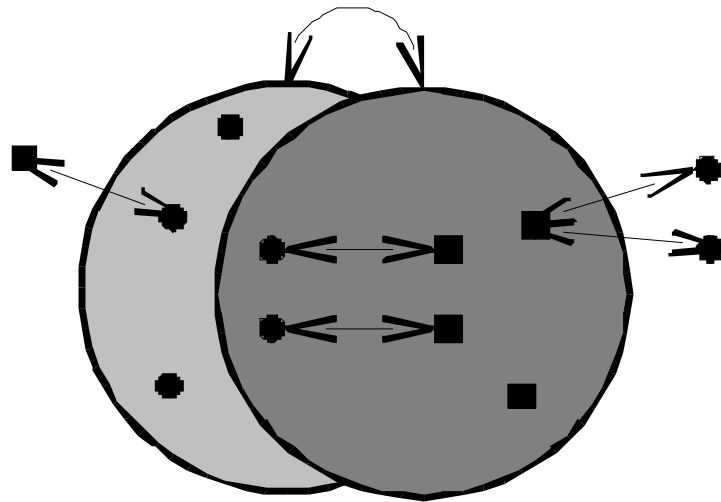


Voss (2006): Collaborative thesaurus tagging the Wikipedia way

<http://arxiv.org/abs/cs.IR/0604036>

[[Mapping categories]]

- * Articles linked via interlanguage links
[[en:Paul_Otlet]], [[en:Paul_Otlet]]...
- * Categories linked via interlanguage links



- * Inter-Wiki mapping with cooccurrences

http://tools.wikimedia.de/~voj/cgi-bin/cocat?wikilang=en&wikifam=.wikipedia.org&cat=Information_science

[[Mapping categories and KOS]]

- * Manually add links to other KOS
- * [[Kategorie:Informetrie]] <->
JITA class **BB** *bibliometric methods*
<http://de.wikipedia.org/wiki/Kategorie:Informetrie>
- * Not very visible
- * Difficult to judge

[[Mapping articles and KOS]]

- * Links to authority file PND (German National Library)

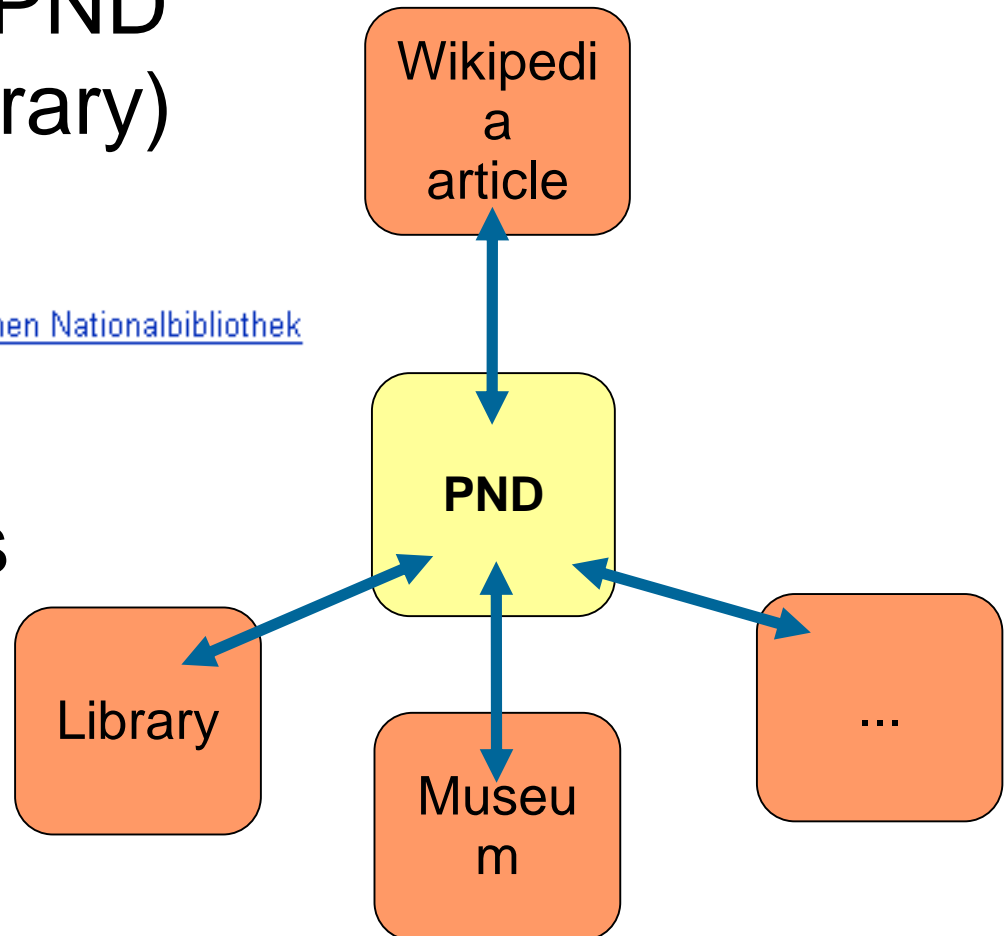
{ { PND | 119510928 } }

■ [Literatur von und über Paul Otlet](#) im Katalog der [Deutschen Nationalbibliothek](#)

- * De, Aug 2006:
>100.000 biographies
~ 20% with PND

- * Plans with GKD (corporate bodies)

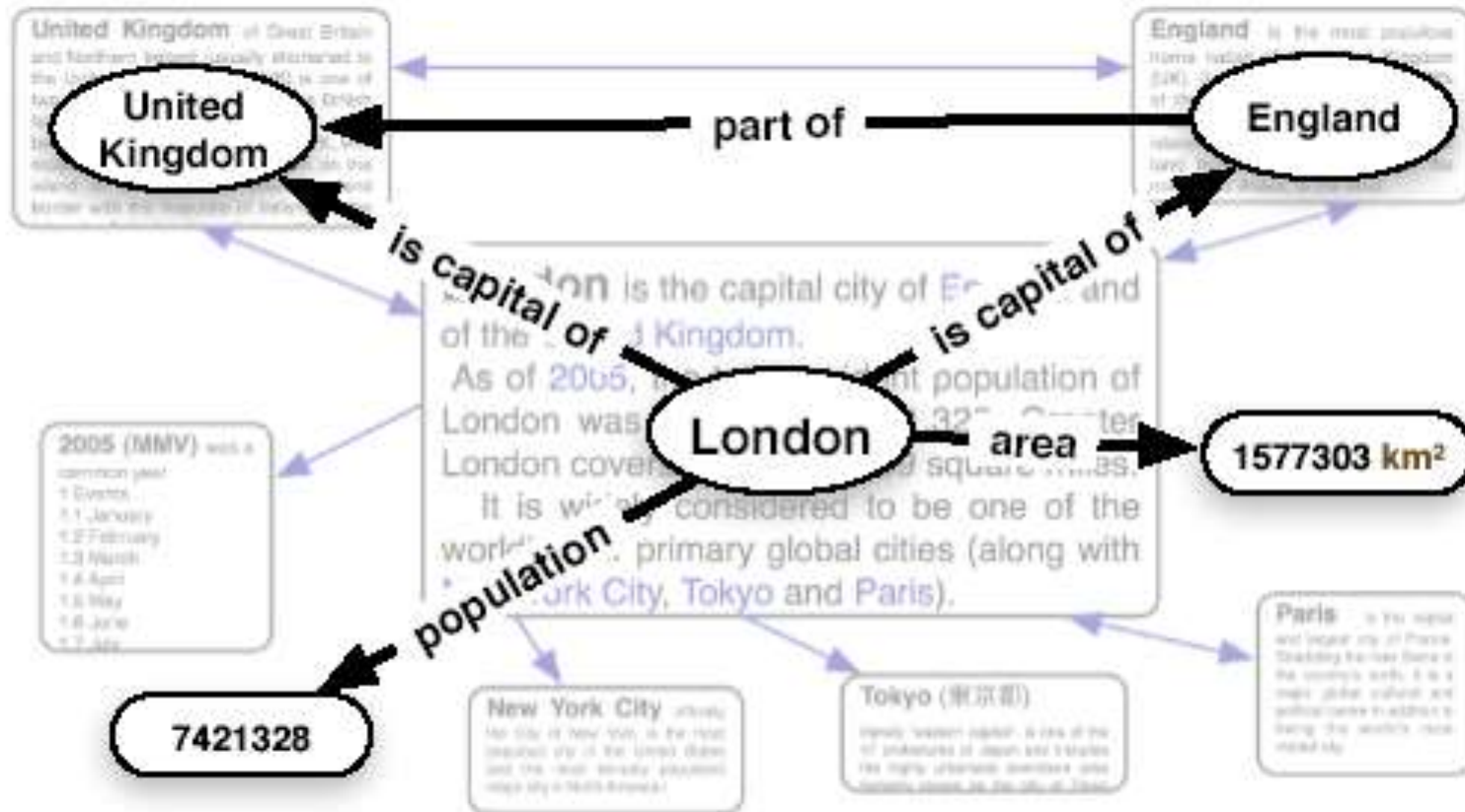
- * 1:1, 1:n, n:m



[[Indexing and the Semantic Web]]

- * Redirects & Disambiguation
=> **Controlled vocabulary**
- * **Stable referable concepts for indexing
(identity and reference in SW)**
- * Proposed by *Bunescu & Pasça (2006), Fernández-García et al (2005), Hepp et al (2006), Krizhanovsky (2006),: Synonym search in Wikipedia: Synarcher, Lüer (2006), Voss (2006)...*
- * **Solution: Published Subject Identifiers**
Pepper (2006)
- * **Changes and stable version URI/URL**

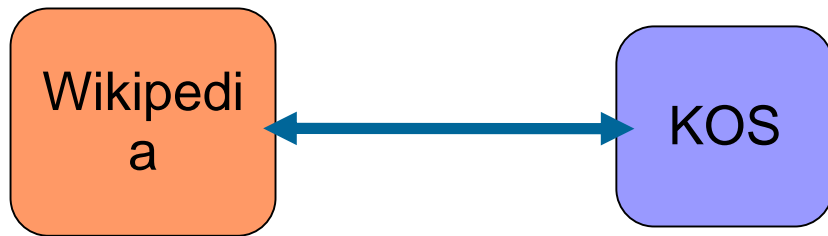
[[Semantic Wikipedia]]



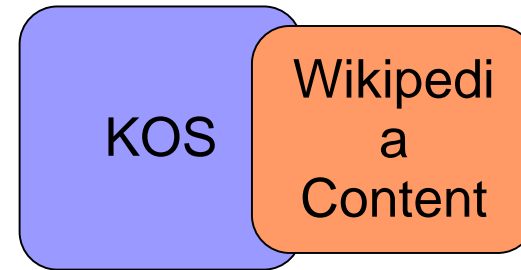
Völkel et al (2006): Semantic Wikipedia. WWW2006 conference

Typed links: `[[is capital of::England]]`
=> RDF triples

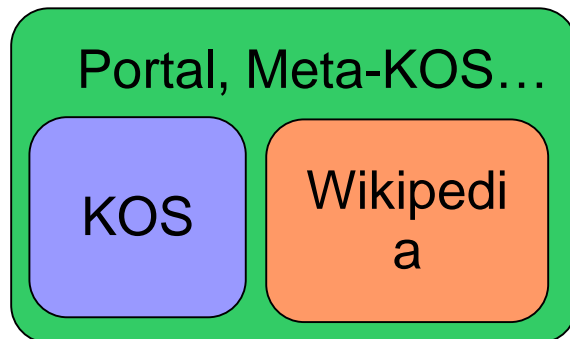
[[Join Wikipedia and KOS]]



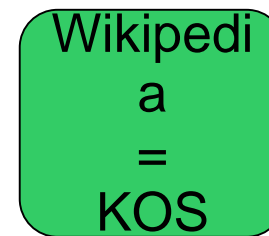
Linking and Mapping



Enrichment



Integration



Combination

[[Learn from **Wikipedia**]]

- * Free you research (open access)
- * Free your data (verifyability)
- * Free your code (open source)
- * Free your KOS (interoperability)

Full references are listed at <http://bibliography.wikimedia.de>