

Interwiki: czy poprawiać, jak poprawiać?

Łukasz Bolikowski

ICM, Uniwersytet Warszawski + IBS, Polska Akademia Nauk

Konferencja Wikimedia Polska 2009, 1-3 maja 2009



Odnosiniki międzyjęzykowe w Wikipedii

Odnosińnik międzyjęzykowy (ang. **interlanguage link**, częściej **interwiki link**) to odnosińnik z artykułu w jednej edycji językowej Wikipedii do odpowiadającego mu artykułu (do artykułu na ten sam temat) w innej edycji językowej.

Innymi słowy, jest to rodzaj **tłumaczenia**.

Szerzej, odnosińniki międzyjęzykowe mogą dotyczyć także kategorii, szablonów, itp.

Dla wygody będę używał krótszych, choć niepoprawnych językowo, terminów:

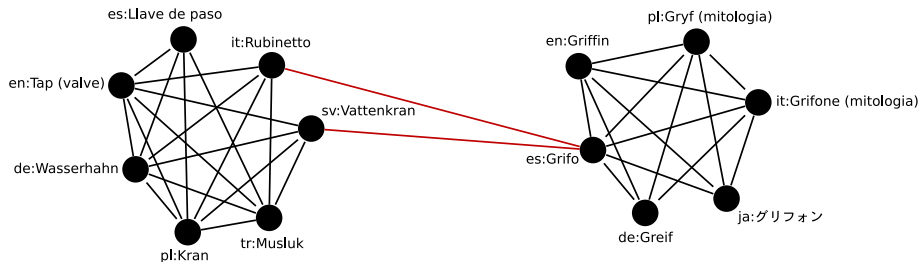
- *odnosińnik* → *link*
- *odnosińnik międzyjęzykowy* → *(link) interwiki*



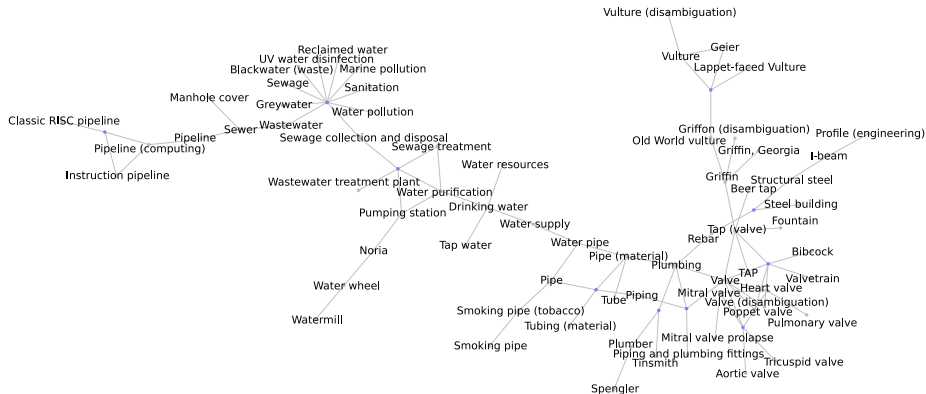
Problemy z interwiki

Silnik MediaWiki nie gwarantuje spójności interwiki – brak centralizacji. Zamiast tego, autonomiczne edycje językowe przechowują „swoje” linki wychodzące. Przy **11,5 mln** artykułów i **90 mln** interwiki prowadzi to do problemów.

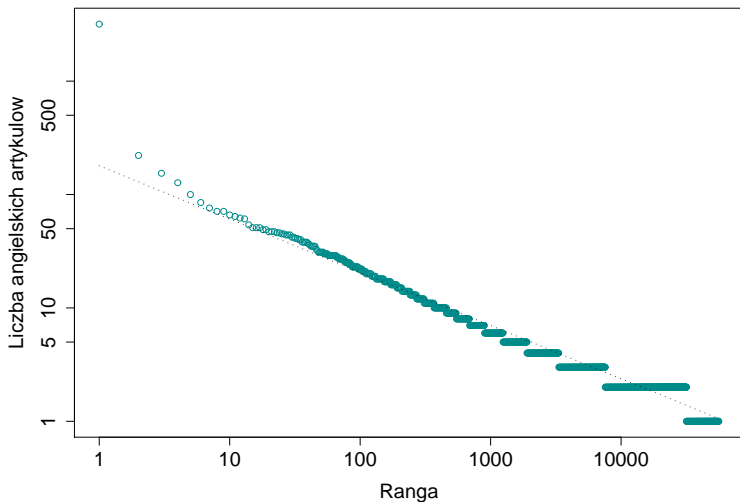
Dwie klasy problemów: brakujące i nadmiarowe linki.



Skala zjawiska



Skala zjawiska

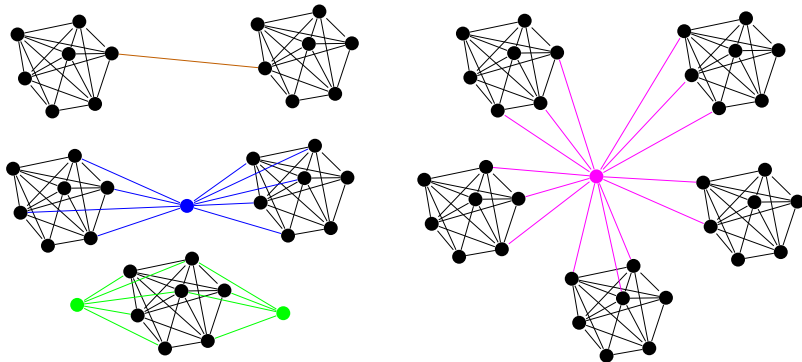


Największa składowa zawiera ok. **80 tys.** artykułów, w tym ponad **3,7 tys.** angielskich (a więc co najmniej tyle różnych tematów), oraz ponad **3,7 mln** linków interwiki. Ręczne przejrzanie wszystkich linków nie wchodzi w grę.

Oprócz tej jest ponad **65 tys.** innych składowych zawierających konflikty.



Najczęściej występujące wzorce



- wandalizmy, absurdalne pomysły
(ro:Nicolae Steinhardt → de:Penis; fr:Rick Ankiel → ja:日本語)
- różne błędy związane z datami, „copy-and-paste”, „off-by-one”
(wuu:5月26号 → bn:??? ← wuu:5月27号; az:3 iyun → su:3 Juli)
(en:10s BC → en:20s BC → en:30s BC → ...)
- linki do stron ujednoznaczających, homonimy
(la:Benedictus (nomen) → en:Benedict; it:Rubinetto → es:Grifo)
- rozszerzanie/zawężanie znaczenia, różna „ziarnistość” edycji językowych
(pl:Województwo krakowskie (I Rzeczpospolita) → en:Kraków Voivodeship (14th century-1795) → pt:Voivodia da Cracóvia → pl:Województwo krakowskie)
- interwiki + redirect
(en:Mother-in-law → ru:Тёща ↔ ru:Родство → en:Kinship)
- różnice kulturowe, ograniczenia tłumaczenia
(en:Pierogi — en:Cepelinaí — en:Dumpling — en:Vareniki — en:Kalduny)

Jak poprawiać?

- 1 szukanie składowych z konfliktami
- 2 szukanie minimalnego zbioru cięć + podział składowej na znaczenia
- 3 prezentacja wyników (wikitools.icm.edu.pl)
- 4 **ocena poprawności podziału składowej na znaczenia**
- 5 **automatyczne usunięcie linków wprowadzających konflikty**



- Serwis WWW: <http://wikitools.icm.edu.pl/>
- Wszystkie konflikty interwiki wraz z rekomendacjami (na podstawie zrzutów bazy z listopada 2008)
- Metoda może być bezpośrednio zastosowana do: Wikisources, Wikibooks, kategorii Wikipedii



Dziękuję

© 2009 Łukasz Bolikowski. Ten utwór jest dostępny na licencji Creative Commons Uznanie autorstwa 2.5 Polska.

Treść licencji dostępna pod adresem: <http://creativecommons.org/licenses/by/2.5/pl/>

