toolserver.org/~lvova

About the project

The aim of Connectivity project is to study and enhance the connectivity, or coherence, of Wikipedia. Or, in other words, to improve hypertext navigation between articles.

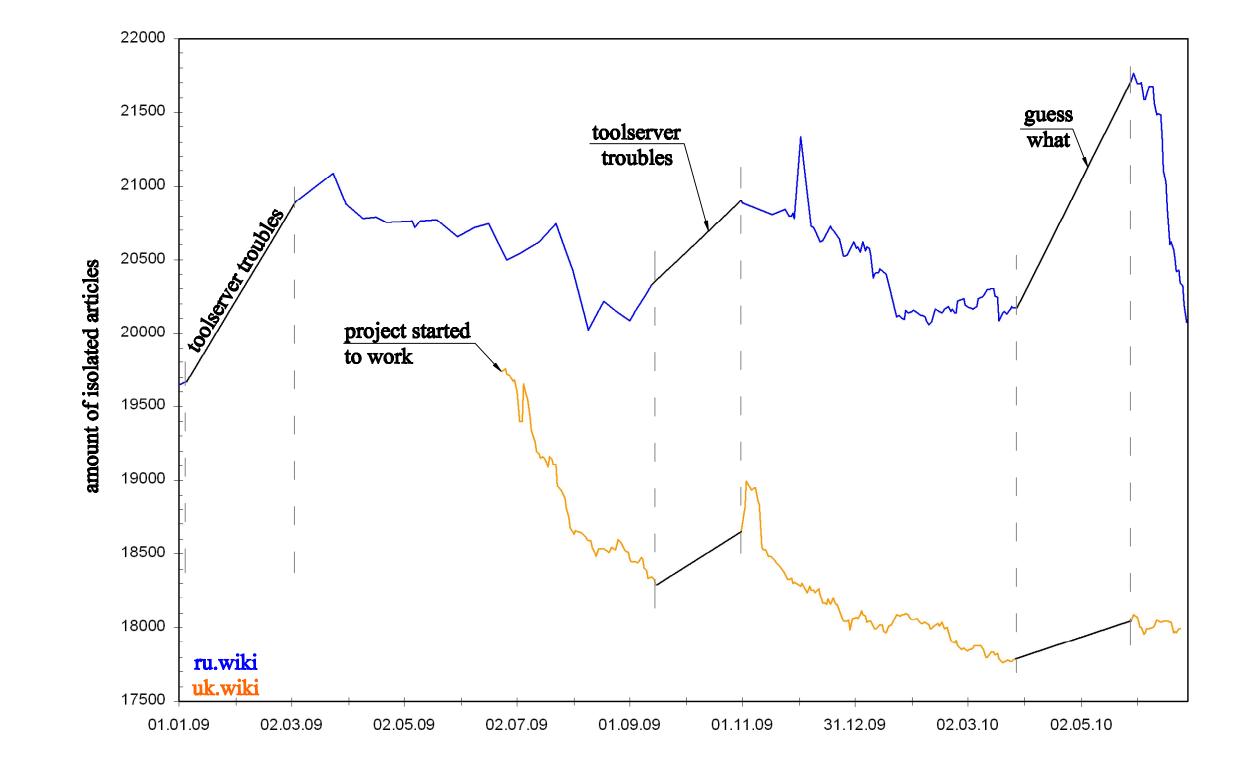
The project deals with deadend pages, isolated articles, non-categorized articles, transitivity of the categorytree and everything in between and

The project works since 2007. Now it has five localisations (de, en, eo, ru, uk) and processes 145 Wikipedias.

Besides identifying connectivity problems, the project also help solving them: there is a Toolserver-based website containing analysis results and linking tips.

Connectivity analysis helps discovering a variety of shortcomings in present articles: lack of notability, errors in naming, lack of necessary redirects, wrong links leading to disambiguation pages and so on. Many isolated articles are in fact subject to deletion or serious modification due to reasons unrelated to connectivity.

Finally, many search engines are known to take the quantity and quality of references to a page into account in their algorithms. Improving connectivity eventually leads to an increase of article's PageRank (as calculated by Google) and therefore to raising articles in search results, thus involving new readers.



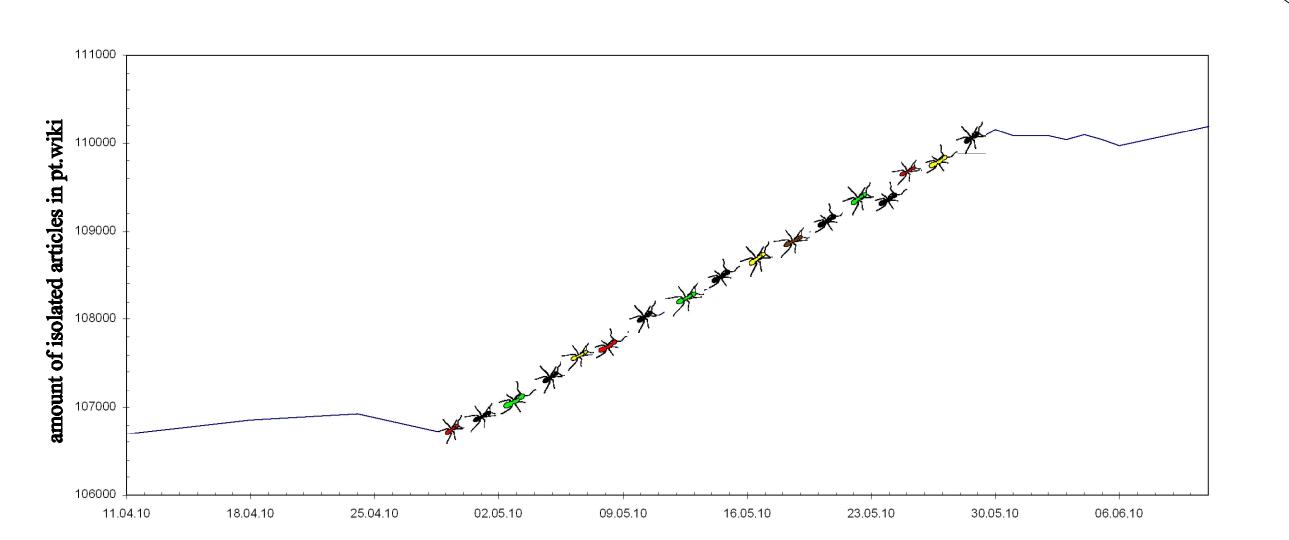
The number of isolated articles is directly related to the authors' awareness of lack of referencing articles. During periods of Toolserver's problems the number of isolated articles usually grows, and the growth gradually turns to decline once the project start to work again.

amount of articles according to Connectivity's definition of article

Isolated articles

– are the ones that cannot be reached through links from the vast majority of other pages. For convenience they're defined as unreachable from the main page. A major difference between isolated articles and the rest of articles with quality notes is that fixing the isolated requires editing other articles.

The most frequent type of isolated articles recognized even by the MediaWiki are the orphaned pages. They are not referenced by other pages of the main namespace. But special pages do not provide with adequate information on connectivity: on-wiki lists are limited to 5000 entries, the accounting methods are imperfect. An orphaned page is defined as an article not referenced by other articles, belonging to the main namespace, except disambigs, coordination lists and calendar articles.



May 2010. Thousands of articles about ants that were uploaded in pt.wiki found themselves in isolation.

We don't know who and how does it in Polish Wikipedia, but we would like to thank him or them.

You're cool!

Moreover, besides orphaned pages that are not referenced, the project sort out isolated pages. These articles may contain links, but, as mentioned previously, it's impossible to reach them from the title page using link navigation. They are orphaned pages that link to each other. In other words, isolated pages are orphaned pages grouped into isolated clusters. Being connected by independent links the clusters may group into chains of isolated clusters.

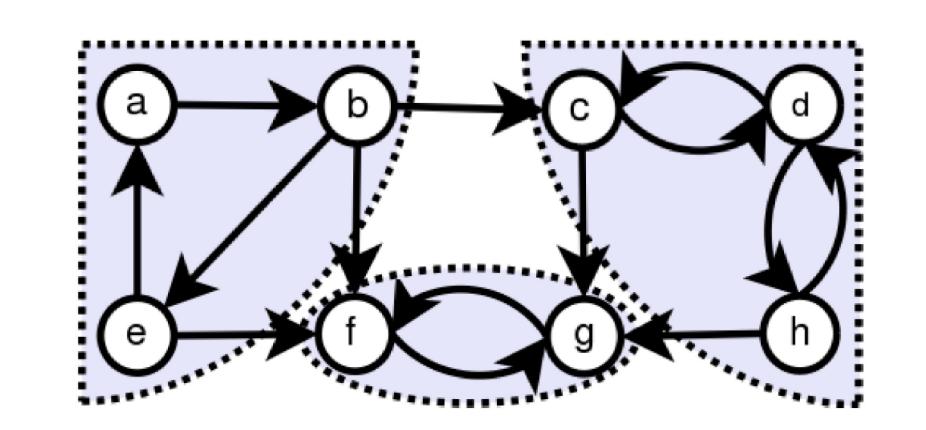
Disambigations pages

– are pages that describe multiple meanings of a term. Collaborative lists and redirection pages contain no connected text, so they all cannot be treated as articles according to the definition above.

01.01.10

Besides working on articles connectivity, there's important improvement of inter-article links' accuracy. Inaccurate links usually appear when authors forget to check, where the link they create leads to, or when a page that was once an article, becomes an index (or a disambig). One of the Connectivity project tools contains a list of pages ordered by number of references to disambigs, as well as a list of disambigs, ordered by number of referencing articles. This list only accounts for incorrect links.

Using these lists and the scripts that can be found on the website, one can quickly "shorten" the link paths providing more accurate links. By the way, the links that reference a disambig-page using {{other-meanings}} template in an article are considered to be correct.



01.06.10

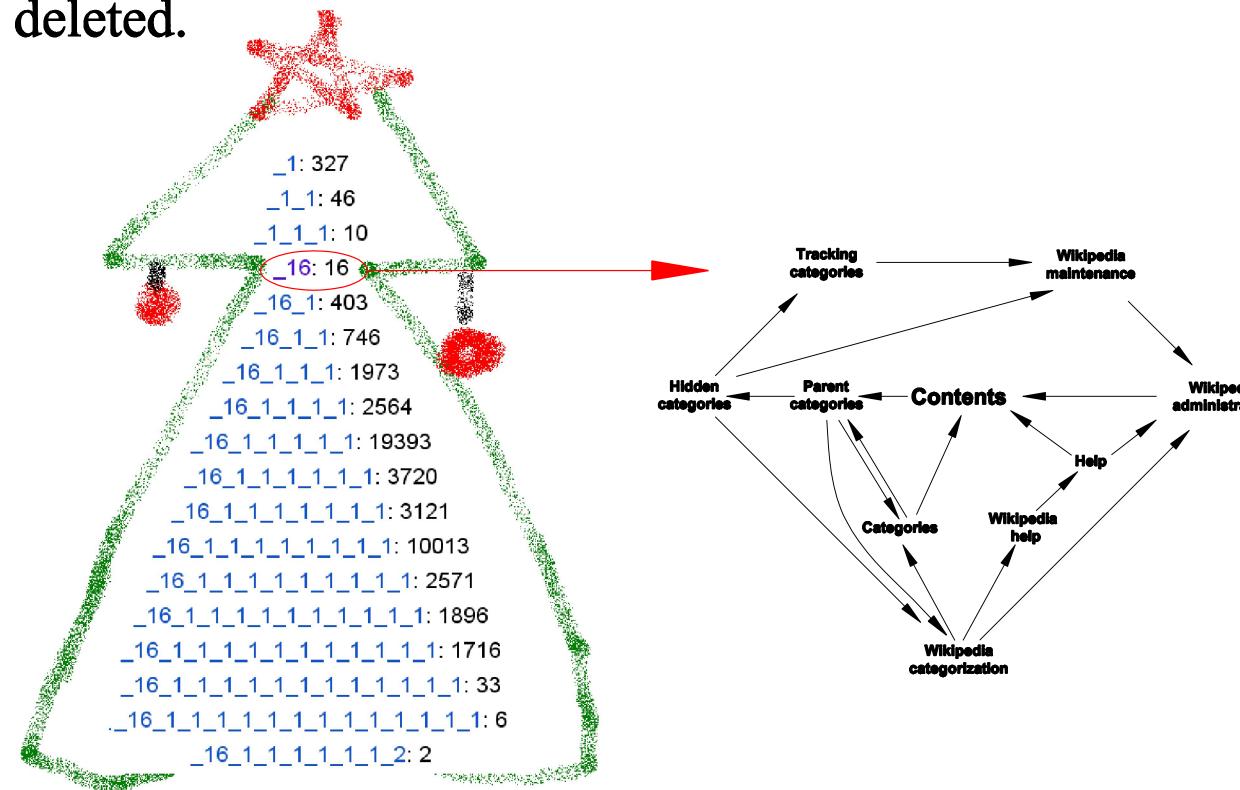
Scientists study thousands of possible forms of clusters and cluster chains. You shouldn't care about the shape. Just link them all.

Categorytree connectivity

Sometimes the categorization transitivity gets disrupted, and categories get linked into cycles of various lengths, which leads to the topological ordering getting thinner. These cycles should be unlinked, restoring the logical ordering thus making readers' navigation over categories easy and pleasant.

Another thing to watch after is keeping the top layer named _1 with just one category. This is the "Contents" Category, called "Everything" in some languages. The other categories from the root layer are simply uncategorized.
Besides all this, the page on the site that

dedicated to categorytree also may contain a list of empty redirecting categories that are to be



This is a part of topological ordering of the English Wikipedia categories graph.

How to set the project up

You are encouraged to spread the Connectivity project by setting it up in your Wikipedia and getting all its benefits.

The project currently can't fully process English Wikipedia, as some routines are too resource-consuming. You are welcome to help optimizing the code!:)

Basic setup, like building category tree or broken redirect lists, is already done for all languages around.

List all templates used to mark disambiguation pages in MediaWiki:Disambiguationspage (sysop rights required to edit). This will enable Connectivity to tell disambiguations from normal pages and to increase statistics' reliability by excluding disambig links from the calculations.

At the time of this writing, 145 wikis already have such a page.

You also may define some types of articles which are not forming valid links: for example, chronological articles or coordination lists.

If you'll translate the website interface to your language, more editors of your wiki will make sense from what's going on.

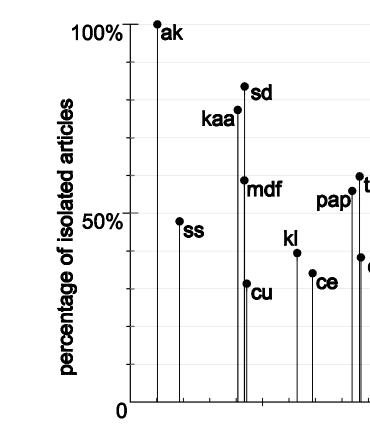
In Russian and Ukrainian wikipedias there are bots using AWB for updating templates on articles that have problems, or used to have them. If your community has a consensus on running such a bot, you may count on project's assistance.

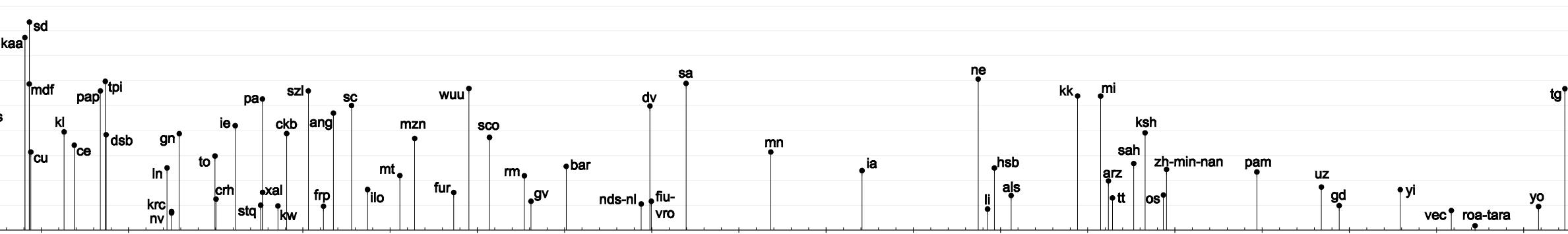
Please take a leaflet and contact the project members later for additional information.

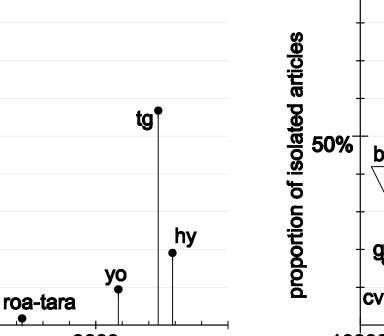
Dead-end pages

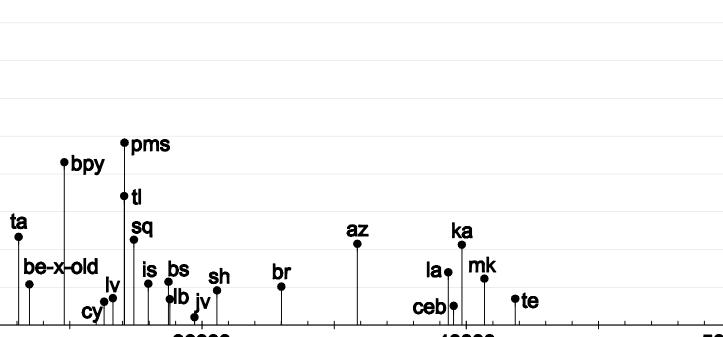
- are the pages of the main namespace lacking links.

MediaWiki's counter is known not to take an article lacking links for a real article, and certainly makes a list of such articles on a special system page. The system list turns out to be inefficient. Connectivity tools enables to form a much longer list of dead-end pages without a limit of 5000, not taking into account links outside the main namespace, links to non-existing pages, links to disambigs, and articles on dates.

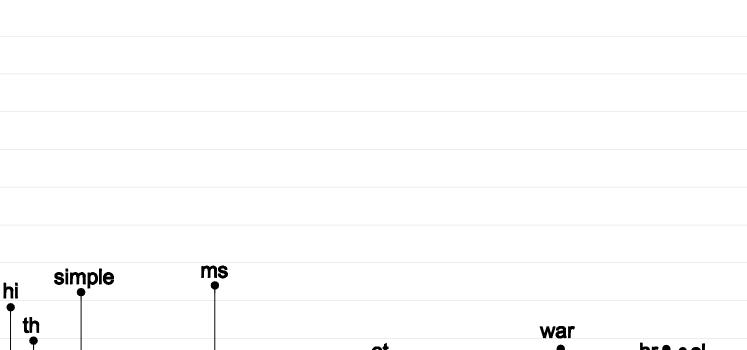




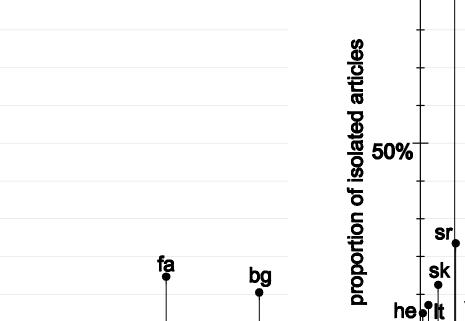


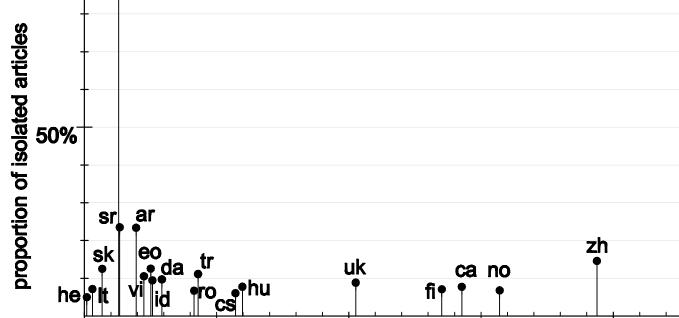


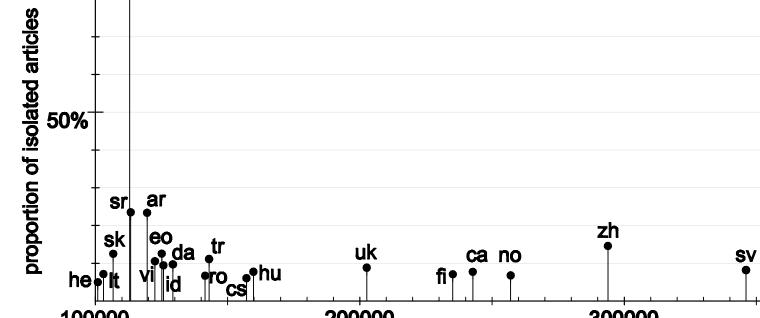
amount of articles according to project's definition of article



02.07.10







amount of articles according to project's definition of article

