PRESENTATION ON yaCy

PRESENTED BY:Nancy vaish
M.Tech (C.S.E) 1st Year



INTRODUCTION

- ➤ YaCy is a peer-to-peer search engine, meaning that there is no centralized authority or server where your information is stored.
- ➤ YaCy pronounced as ("Ya see").
- ➤Its core is a computer program written in Java distributed on several hundred computers so-called YaCy-peers. Each YaCy-peer independently crawls through the Internet, analyzes and indexes found web pages, and stores indexing results in a common database which is shared with other YaCy-peers using principles of P2P networks.
- ➤ YaCy is available on Windows, Mac and Linux.

COMPONENTS

YaCy search engine is based on four elements:

- **≻**Crawler
- **≻**Indexer
- ➤ Search and Administration interface
- ➤ Data Storage

yaCy SEARCH PAGE



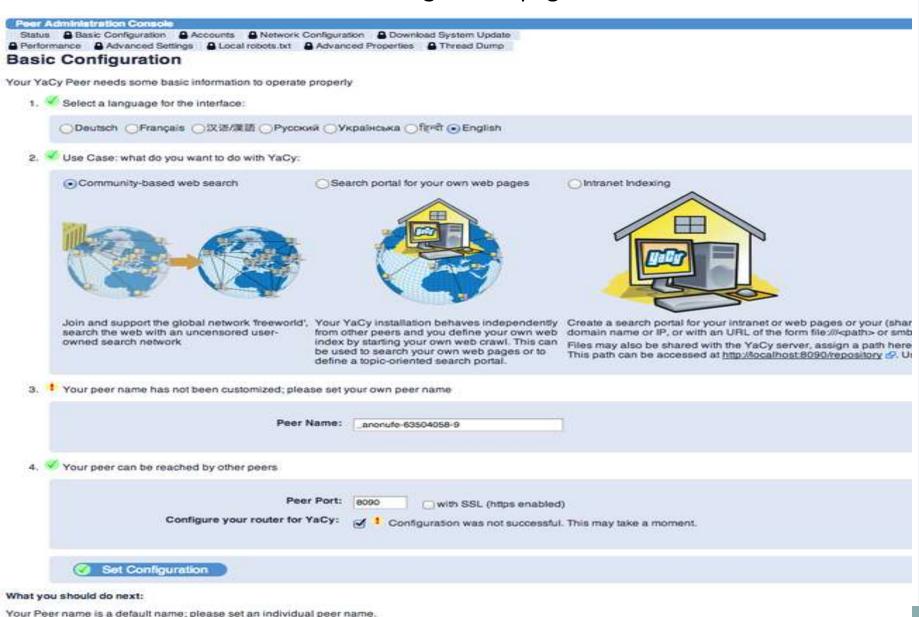
Web Search by the People, for the People

		Search
●Text Images	more options	

As you can see, this is a pretty conventional search engine page. You can search using the provided search bar without any additional configuration, if you wish.

We will be exploring the administration interface though, because that provides us with a lot more flexibility. Click on the "Administration" link in the upper-left corner of the page:

You will be taken to the basic configuration page:



Crawl Sites to Contribute to the Global Index

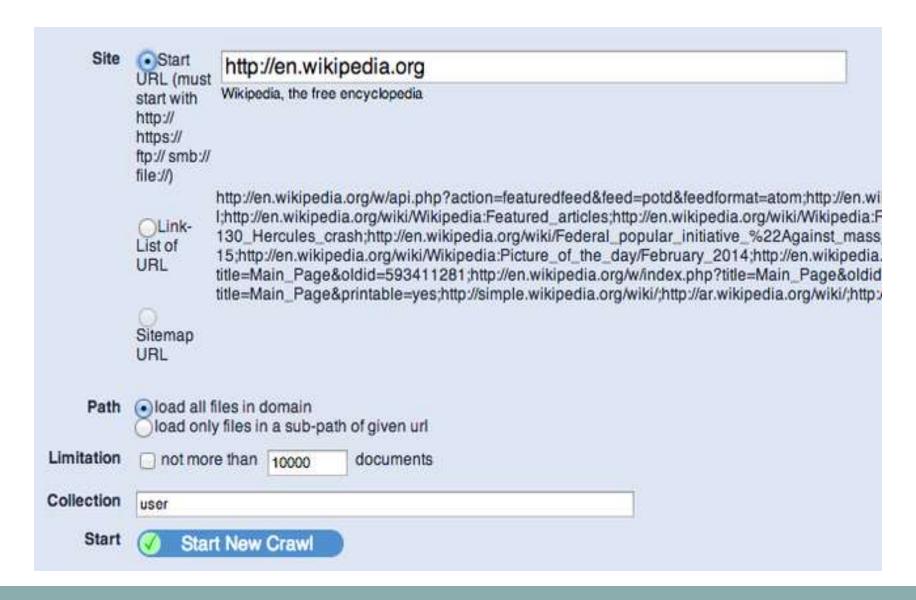
You can now search using the indexes kept on your YaCy peers. The search results will become more and more accurate the more people participate in the system.

To start this process, click on the "Crawler / Harvester" link under the "Index Production" section.

Index Production		
8	Crawler / Harvester	
8	Creation Monitor	
8	Index Administration	
8	Filter & Blacklists	
8	Content Semantic	
8	Target Analysis	
	Process Scheduler	

If you've attempted to search for something and did not get the results you were looking for, consider starting to index the pages on a site with your instance. It will make your search more accurate for yourself and your peers.

Type in the URL that you want to index in the "Start URL" section:



You can limit the number of documents that your crawl will index. Click "Start New Crawl" when you are finished to begin crawling the selected site. Click on the "Creation Monitor" link on the left-hand side to see the progress of the indexing. You should see something like this:



Your server will crawl the URL specified at the rate of 2 requests per second until it has either run out of links chained together or reached the limit you set.

ADVANTAGES

- As there is no central server, the reliability is higher, because there's no single point of failure.
- ➤ Because the engine is not owned by a company, there is no centralized advertising.
- ➤ It is possible to achieve a high degree of privacy.
- ➤On every search YaCy fetches the pages provided in search results and verifies that they still contain the keywords requested by the user. This ensures that the pages that no longer contain the requested keywords are not displayed to the user

DISADVANTAGES

- As there is no central server and the YaCy network is open to anyone, malicious peers are able to insert inaccurate or commercially biased search results.
- Result verification is done client-side on every search, which increases network traffic on the computer running YaCy and makes YaCy slower to display the search results than search engines such as Google.
- ➤ Missing IPv6 support.

CONCLUSION

If you need a great search engine for your site, YaCy provides that option as well. YaCy is very flexible and is an interesting solution to the problem of privacy concerns.

THANK YOU

