

Clustering web search results

Egon Willighagen 

Published April 21, 2007

Citation

Willighagen, E. (2007). Clustering web search results. In *chem-bla-ics*. chem-bla-ics. <https://doi.org/10.59350/ytpss-fw666>

Keywords

Google, Cheminf

Abstract

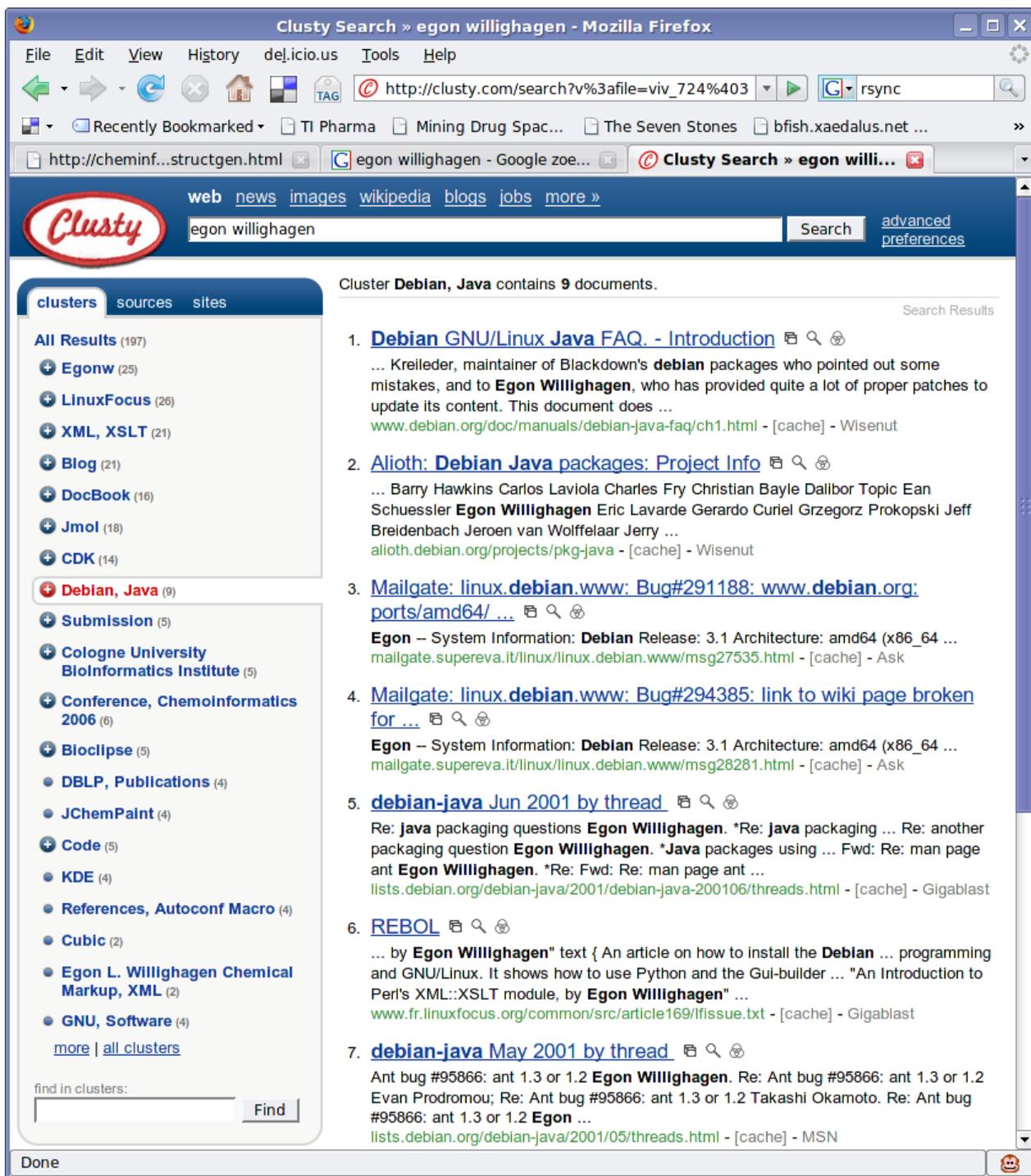
The Dutch Intermediair magazine of this week had a letter sent by a reader introducing Clusty, a web search engine that clusters the results.

Copyright

Copyright © Egon Willighagen 2007. Distributed under the terms of the [Creative Commons Attribution 4.0 International License](#), which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

chem-bla-ics

The Dutch [Intermediar](#) magazine of this week had a letter sent by a reader introducing [Clusty](#), a web search engine that clusters the results. It does a pretty good job for 'egon willighagen':



Clusty Search » egon willighagen - Mozilla Firefox

File Edit View History del.icio.us Tools Help

http://clusty.com/search?v%3afile=viv_724%403 rsync

Recently Bookmarked TI Pharma Mining Drug Spac... The Seven Stones bfish.xaedalus.net ...

http://cheminf...structgen.html egon willighagen - Google zoe... Clusty Search » egon willi...

web news images wikipedia blogs jobs more »

egon willighagen Search advanced preferences

clusters sources sites

All Results (197)

- Egonw (25)
- LinuxFocus (26)
- XML, XSLT (21)
- Blog (21)
- DocBook (16)
- Jmol (18)
- CDK (14)
- Debian, Java (9)**
- Submission (5)
- Cologne University Bioinformatics Institute (5)
- Conference, Chemoinformatics 2006 (6)
- Bioclipse (5)
- DBLP, Publications (4)
- JChemPaint (4)
- Code (5)
- KDE (4)
- References, Autoconf Macro (4)
- Cubic (2)
- Egon L. Willighagen Chemical Markup, XML (2)
- GNU, Software (4)

more | all clusters

find in clusters: Find

Cluster **Debian, Java** contains 9 documents.

Search Results

- [Debian GNU/Linux Java FAQ. - Introduction](#) ... Krelleder, maintainer of Blackdown's **debian** packages who pointed out some mistakes, and to **Egon Willighagen**, who has provided quite a lot of proper patches to update its content. This document does ...
www.debian.org/doc/manuals/debian-java-faq/ch1.html - [cache] - Wisenut
- [Alioth: Debian Java packages: Project Info](#) ... Barry Hawkins Carlos Laviola Charles Fry Christian Bayle Dalibor Topic Ean Schuessler **Egon Willighagen** Eric Lavarde Gerardo Curiel Grzegorz Prokopski Jeff Breidenbach Jeroen van Wolfelaar Jerry ...
alioth.debian.org/projects/pkg-java - [cache] - Wisenut
- [Mailgate: linux.debian.www: Bug#291188: www.debian.org: ports/amd64/ ...](#)
Egon -- System Information: **Debian** Release: 3.1 Architecture: amd64 (x86_64 ...
mailgate.supereva.it/linux/linux.debian.www/msg27535.html - [cache] - Ask
- [Mailgate: linux.debian.www: Bug#294385: link to wiki page broken for ...](#)
Egon -- System Information: **Debian** Release: 3.1 Architecture: amd64 (x86_64 ...
mailgate.supereva.it/linux/linux.debian.www/msg28281.html - [cache] - Ask
- [debian-java Jun 2001 by thread](#)
Re: **java** packaging questions **Egon Willighagen**. *Re: **java** packaging ... Re: another packaging question **Egon Willighagen**. ***Java** packages using ... Fwd: Re: man page ant **Egon Willighagen**. *Re: Fwd: Re: man page ant ...
lists.debian.org/debian-java/2001/debian-java-200106/threads.html - [cache] - Gigablast
- [REBOL](#)
... by **Egon Willighagen** text { An article on how to install the **Debian** ... programming and GNU/Linux. It shows how to use Python and the Gui-builder ... "An Introduction to Perl's XML::XSLT module, by **Egon Willighagen**" ...
www.fr.linuxfocus.org/common/src/article169/ffissue.txt - [cache] - Gigablast
- [debian-java May 2001 by thread](#)
Ant bug #95866: ant 1.3 or 1.2 **Egon Willighagen**. Re: Ant bug #95866: ant 1.3 or 1.2 Evan Prodromou; Re: Ant bug #95866: ant 1.3 or 1.2 Takashi Okamoto. Re: Ant bug #95866: ant 1.3 or 1.2 **Egon** ...
lists.debian.org/debian-java/2001/05/threads.html - [cache] - MSN

It seems to use other engine to do the searching and focus on the clustering. Source engine exclude Google, and include [Gigablast](#), [MSN](#) and [Wikipedia](#).

For *chemoinformatics* it comes up with the following top 10 clusters: 'Drug Discovery', 'Structure', 'Cheminformatics', 'Research', 'Books', 'Conference, German', 'Textbook, Gasteiger', 'Laboratory', 'Handbook of Chemoinformatics', and 'School'. Quite acceptable and useful clustering.

chem-bla-ics

This might be the next step in googling. Rich, it also might solve [your problem](#) : searching for 'ruby chemoinformatics' does **not** give a 'Depth First' or 'Rich Apodaca' cluster :)