

# Volomonitor - the Internet Gold Miner

Anthony Wood

One of the most powerful information retrieval techniques is known as Pearl Growing. Just as an oyster can take a speck of sand and turn it into a pearl, a searcher can use the results of a simple search to create a highly relevant set of documents.

Commercial online news services do this by attaching *More Like This* commands to retrieved articles. Clicking on the command initiates a second search that uses the index terms / tags attached to the article to find similarly indexed articles. The results are not usually that good unless the indexing is extensive as there is too little text, i.e. too few clues, to work on.

Good information researchers can manually analyse a number of articles that closely match the desired specification and construct a second search using a mixture of index terms and words from the headlines and body text. This, however, can take a lot of time.

So much for commercial online services what about the Internet?

The problem with the Internet is that it does not have the structure of a commercial online database where sources are hand-picked and each document is carefully indexed for geographical location, industry sector and subject. There are huge variations from one website to another. However, the Internet is too rich a source of information to be ignored when gathering data.

One company that has "tamed" the Internet is Volocom from Italy. They use a variation of the Pearl Growing two-stage search with Google for the first stage and a very sophisticated data-mining application, Concept Searching, for stage two.



Users first create a Dossier to hold both the raw data "ore" and the articles or "nuggets" selected by Concept Searching's filtering / enrichment processes. The former are known as Imported Documents and the latter Approved Documents. The end result is illustrated in the figure.

The dossier I created was called Fuel Cells, 961 articles were imported for analysis and 5 of these were selected for retention as Approved documents leaving 961 for further analysis. This is the endpoint. What follows is how I got there.

First a brief word about fuel cells. They convert hydrogen directly into electricity and water with near 100% efficiency. They are, therefore, a prime candidate for an environmentally-friendly means of generating electricity and powering cars, trucks, buses and trains. Discovered in 1839, they are a true "Cinderella" technology. They have been used extensively in NASA space missions to provide the hotel load aboard spacecraft. Interest in terrestrial applications has mushroomed in the last 10 years and they are on track largely to replace the very wasteful internal combustion engine by the middle of the 21<sup>st</sup> century.

The first stage in a Volomonitor search is to gather relevant documents from the Web using Google. Google has rapidly established itself as the dominant Web search engine and my own experience of using it has been very positive. Google, however, makes its money by promoting websites that sell products and services so some of the highest ranked hits may have got there more from their owners' ability to pay rather than their relevance. This apart it is a very powerful retrieval tool.

On the page opposite is the Import Preview page that is displayed following the entry of search text into the Volomonitor search screen.

The search text was *fuel cell*, the maximum number of retrieved documents was set at 1,000 and English only was the language choice. As can be seen from the above Google found 9.6 million pages with *fuel* and / or *cell* in them. The top 25 relevance-ranked articles are presented as a final check prior to starting the Import process in earnest. The section of the Dossier – Import history screen below details what happened next.

The import process took just under an hour i.e. from 15:40 to 16:37. 966 English-language articles were retrieved and every word in every article was indexed by Concept Searching. This is not the "smash and grab" process of a conventional Web search – good results require the same patience as that of the fisherman. This is not a system for those who need a quick fix.

It is possible to add to the dossier by clicking on the Run again button. In this way an increasing collection of articles on a specific subject area can be isolated and processed for interrogation using Concept Searching's analytical tools. One immediate benefit of the creation of this mini-database is a listing of the Top sources / websites used and the numbers of articles retrieved from them. This can be a route to adding useful websites to a favourites list.

olomonitor

User: Anthony Wood  
Powered by Volocom  
Powered by conceptSearching

Home page | My dossiers | Advanced search | My folders | Help | Exit

**My dossiers** ? | **Dossier - Import**

**Import preview**

This is the preview of your import. Please check the first resulting documents, then if correct click "Start import".

Modify search Start import

Search text: **fuel cell**  
Max. doc number: **1000**  
Language: **English**

Web Immagini Gruppi Directory News altro »

Google fuel cell Cerca Ricerca avanzata Preference

Cerca:  il Web  pagine in italiano  pagine provenienti da: Italia  Modifica

**Web** Risultati: 1 - 25 su circa 9.570.000 pagine in Inglese per fuel cell. (0,24 secondi)

**Fuel Cells 2000** - [ Traduci questa pagina ]  
Fuel Cells 2000 is a non-profit project which provides educational information on fuel cells to the public, the media, and policy makers.  
[www.fuelcells.org/](http://www.fuelcells.org/) - 15k - 8 ago 2005 - Copia cache - Pagine simili

**Howstuffworks "How Fuel Cells Work"** - [ Traduci questa pagina ]  
According to many experts, we may soon find ourselves using fuel cells to generate electrical power for all sorts of devices we use every day.  
[science.howstuffworks.com/fuel-cell.htm](http://science.howstuffworks.com/fuel-cell.htm) - 37k - 8 ago 2005 - Copia cache - Pagine simili

Collegamenti sponsorizzati

**GEFC Fuel Cells**  
GEFC - 11N PEM (N = 1- 20 mil)  
World Leading PEM, CCM, MEA, MFC  
[WWW.GEFC.COM](http://WWW.GEFC.COM)

**HyTec**  
Master of Science  
Fuel Cells and Hydrogen Technology  
[www.hytec.aau.dk](http://www.hytec.aau.dk)

Open in new window

All materials published on the Web site are protected in any manner without the express written permission of the publisher.

**Dossier - Import history**

**All imports**

- fuel cell**
  - From the Web (Google)
  - Max. doc number: 1000
  - Language: en
  - Runs: 08/08/2005 h15:40 Completed il 08/08/2005 h16:37
  - From this import: 966 documents.

Run again Hide

The next stage is to interrogate the fully-indexed Fuel Cells mini-database. Searching using the term *buses* to retrieve articles on fuel cell buses retrieved 139 of the 966 articles as can be seen from the screen below.

Volocom - Dossier - Imported documents - Microsoft Internet Explorer

File Edit View Favorites Tools Help

Back Forward Stop Search Favorites Media

Address: <http://volomonitor.volocom.it/eng/page.asp>

Copernic Agent The Web Up History Track Results Bar

RobForm Search Logins Volocom (+1) Frank Ryan Save Generate

olomonitor

User: Anthony Wood  
Powered by Volocom  
Powered by conceptSearching

Home page | My dossiers | Advanced search | My folders | Help | Exit

**My dossiers** ? | **Dossier - Imported documents**

**Imported documents** (966)

Search inside this dossier:

Language: English

From date:

To date:

Precision: Max

Source: All

Go!

Note: it will take some minutes till documents will be available for search process.

**Search result**

Precision: Max High Medium Low Min  
Group in a cluster with affinity Low Medium High Do not group

Page: 1 2 3 4 5 6 >>

139 available documents.

- 08/08/2005 London Buses - Fuel cell buses - Transport for London (<http://www.tfl.gov.uk>)
- 07/06/2005 Clean Air Initiative: Infopool - Fuel Cell Bus Programs Worldwide (<http://www.cleanairnet.org>)
- 08/07/2004 Planned Fuel Cell Bus Demonstrations (<http://www.nvco.org>)
- 09/06/2005 fuel-cell-buses.pdf (<http://www.tfl.gov.uk>)
- 17/02/2004 Fuel Cell Bus Demonstrations, page one (<http://www.nvco.org>)
- 28/01/2005 CD117a-030129.pdf (<http://www.technology.gov>)
- 27/07/2005 Federal Transit Administration - Bus Technologies and Testing Section (<http://www.fta.dot.gov>)
- 17/02/2004 The Fuel Cell (<http://www.fta.dot.gov>)
- 08/08/2005 Green Car Congress: Fuel Cells (<http://www.greencongress.com>)
- 08/08/2005 Ballard Secures Contract for Three Fuel Cell Bus Engines for Australia (<http://www.theautochannel.com>)
- 10/03/2005 NFBCTI background.pdf (<http://www.calstart.org>)
- 08/08/2005 NFBCTI Technology Validation: Hydrogen Fuel Cell Bus Evaluation for California Transit Agencies (<http://www.cars.conrcc.com>)
- 08/08/2005 Fuel Cell Buses, Driving Now - Gizmodo (<http://www.gizmodo.com>)
- 08/08/2005 Georgetown University Advanced Vehicle Development (<http://fuelcellbus.georgetown.edu>)
- 08/08/2005 Fuel cell: Definition and Much More From Answers.com (<http://www.answers.com>)
- 08/08/2005 Green Car Congress: DaimlerChrysler Delivers 3 Fuel-Cell Citare Buses to Beijing (<http://www.greencongress.com>)
- 08/08/2005 Getting on board the fuel-cell bus (<http://www.msnbc.com>)
- 23/12/2004 Links (<http://www.iqas.org>)
- 08/08/2005 Fuel cell - Wikipedia, the free encyclopedia (<http://en.wikipedia.org>)
- 31/07/2005 Fuel cell - Wikipedia, the free encyclopedia (<http://en.wikipedia.org>)
- 24/03/2002 km04r0.pdf (<http://www.ecrc.energy.gov>)
- 04/07/2005 Roadmap.pdf (<http://www.fuelcell-conrad.org>)
- 22/01/2004 The Future of Fuel Cells...solid, oxide, polymer, electrolyte, direct, methanol, phosphoric, acid, alkaline, molten, carbonate, alcohol, natural, gas, automobile, bus, catalyst, carnot, chp, cogeneration, efficiency, electric, gasoline, hydrogen, monoxide, (<http://www.benwiens.com>)
- 08/08/2005 Fuel Cells News, fuelcellworks (<http://www.fuelcellworks.com>)
- 07/12/2004 fuelcell.pdf (<http://www.engr.csc.ku.edu>)

Page: 1 2 3 4 5 6 >>

139 available documents.  
Documents per page: 25 50 100

**History**

**Related topics**

- fuel cell buses
- natural gas
- route 25
- urban transport
- bus technology
- fuel cell bus
- hydrogen fuel cells
- bus demonstration
- fuel cell engines
- transit buses

Asignati alla ricerca i termini selezionati

**Top Source**

A list of the most frequent sources in imported documents.

**London Buses - Fuel cell buses - Transport for London**

The fuel cell buses first went into service in January 2004 on route 25, with two buses in operation each day. The buses ran in addition to the normal service, as London Buses needed to assess how reliable they would be. The buses proved so reliable that when they were switched to RUs, all three buses were put into service and ran as part of the normal timetable.

(<http://www.tfl.gov.uk/dai/08/08/2005>)

- <http://www.fuelcellstore.com> (10)
- <http://www.arconcarconrccs.com> (10)
- <http://www.humboldt.edu> (10)
- <http://www.umib.com> (10)
- <http://www.nextmaterials.com> (10)
- <http://www.rni.org> (10)
- <http://www.theautochannel.com> (10)
- <http://www.tnma.com> (10)
- <http://www.wired.com> (10)

olocom technology

<http://volomonitor.volocom.it/eng/page.asp?page=content&ref=VD-06976&terms=buses,&languages=en>

This is a VERY IMPORTANT screen that illustrates many of the tools that make Volomonitor's use of Concept Searching both unique and revolutionary. Hovering over a retrieved headline displays an extract from the article containing the greatest concentration of search terms – a brilliant way to determine relevance. Of greater importance, however, is the list of Related Topics. These are fuzzy multi-word concepts that have been extracted from the 139 retrieved articles. This feature is unique to Concept Searching and makes it far more useful than many of the "heavyweight", and heavy price, search engines that are in common use.

Related topics can be added to an existing search to sharpen it even further. Checking the boxes for three of these terms produced the following search strategy

***buses + fuel cell buses + fuel cell bus + transit buses***

This resulted in a smaller set of 47 articles. By using the hover technique I was able to choose 5 articles that homed in on exactly the data I needed. These I transferred to my Approved folder for ready reference.

In this evaluation I have not looked at all the functionality and features of Volomonitor. Instead I have homed in on what I consider to be the major benefit of the system.

## Conclusions

I was familiar with the Concept Searching application prior to testing Volomonitor. The usual way of using it is to point it at specific websites of known relevance. What Volomonitor has done is to turn the whole of the Internet into a ready source of information for serious researchers. The power of a conventional Internet search engine is used to extract a manageable number of relevant documents i.e. to create a mini-database. This relatively unstructured collection is then turned into a more structured format using Concept Searching's indexing and multi-word concept extraction facilities. It then has all the attributes of a commercial database including ease of use.

The main value of any information system or service is to save time. If knowledge workers can retrieve information quickly they can spend more time doing what they are paid to do i.e. add value in the form of insights and recommendations for action that bring benefits to the organisations that employ them.

Volomonitor is a highly innovative and unusual information management tool. It is unique. It may well be a harbinger of similar products that "sample" the Internet and then structure the retrieved data to create exclusive views of a particular subject area.

Though Volomonitor is way down the track there are areas for enhancement. Google is good but Google + Yahoo + Altavista + Hotbot + others, a la Copernic, might be better. Also a "Deep Web" search engine, such as BrightPlanet's Deep Query Manager, might reach parts of the Internet that these crawler-type engines do not reach.



## Contact -

### Volocom srl

Via G. da Udine, 34  
20156 Milano

Tel: +39 (0) 238093741

[www.volocom.it/](http://www.volocom.it/)

### Concept Searching Limited

9 Shephall Lane  
Stevenage  
Herts SG2 8DH  
UK

Tel: 44 (0) 1438 213545

[info-uk@conceptsearching.com](mailto:info-uk@conceptsearching.com)

[www.conceptsearching.com](http://www.conceptsearching.com)