Solving the Boolean NOT search question

By Gordon Rugg

Background: This article shows how Search Visualiser can find "not the usual suspects" records, by sidestepping the notorious Boolean "NOT" problem.

This article was originally posted on April 23, 2012

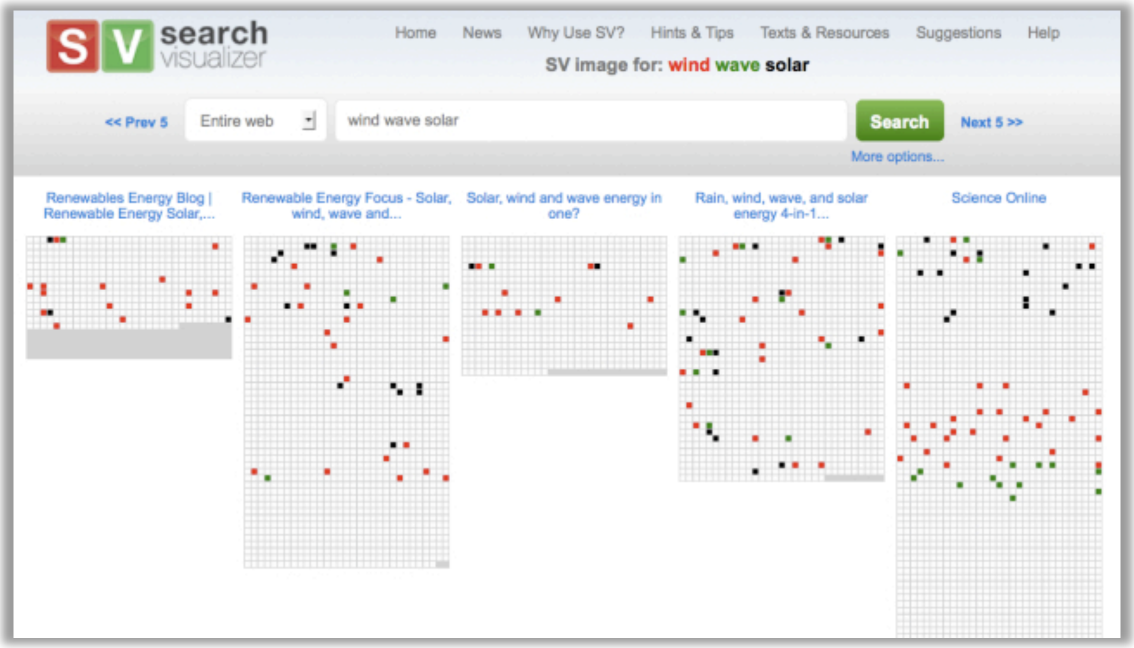
A classic problem in online searching is finding documents which aren't about the usual suspects.

Suppose, for instance, you're trying to find documents about types of sustainable energy other than wind, wave or solar power. The classic novice's reaction is to use an advanced search phrasing along the lines of *sustainable energy NOT wind wave solar*

The problem with this phrasing, as every information retrieval specialist knows all too well, is that it will only find documents which mention sustainable energy and which do not mention the words *wind*, *wave* or *solar*. However, the vast majority of relevant records will mention those words as part of their introduction. Those records would all be discarded because of the NOT, even though they were relevant.

The other obvious solution, namely specifying the "not the usual suspects" energy sources that you want, also hits a major problem, since by definition you'll probably be doing this search precisely because you don't know what those other energy sources are, and you're trying to find out.

One new solution to this problem involves making use of the human ability to make sense of patterns rapidly and efficiently. Here's an example. It's a search for wind wave solar on Search Visualiser. It shows records returned from a standard Bing search. The Search Visualiser has shown where each keyword occurs throughout each record.


S V search visualizer
Home
News
Why Use SV?
Hints & Tips
Texts & Resources
Suggestions
Help
<< Prev 5
Search
Next 5 >>
More options...
Renewables Energy Blog | Renewable Energy Solar,...
Renewable Energy Focus - Solar, wind, wave and...
Solar, wind and wave energy in one?
Rain, wind, wave, and solar energy 4-in-1...
Science Online

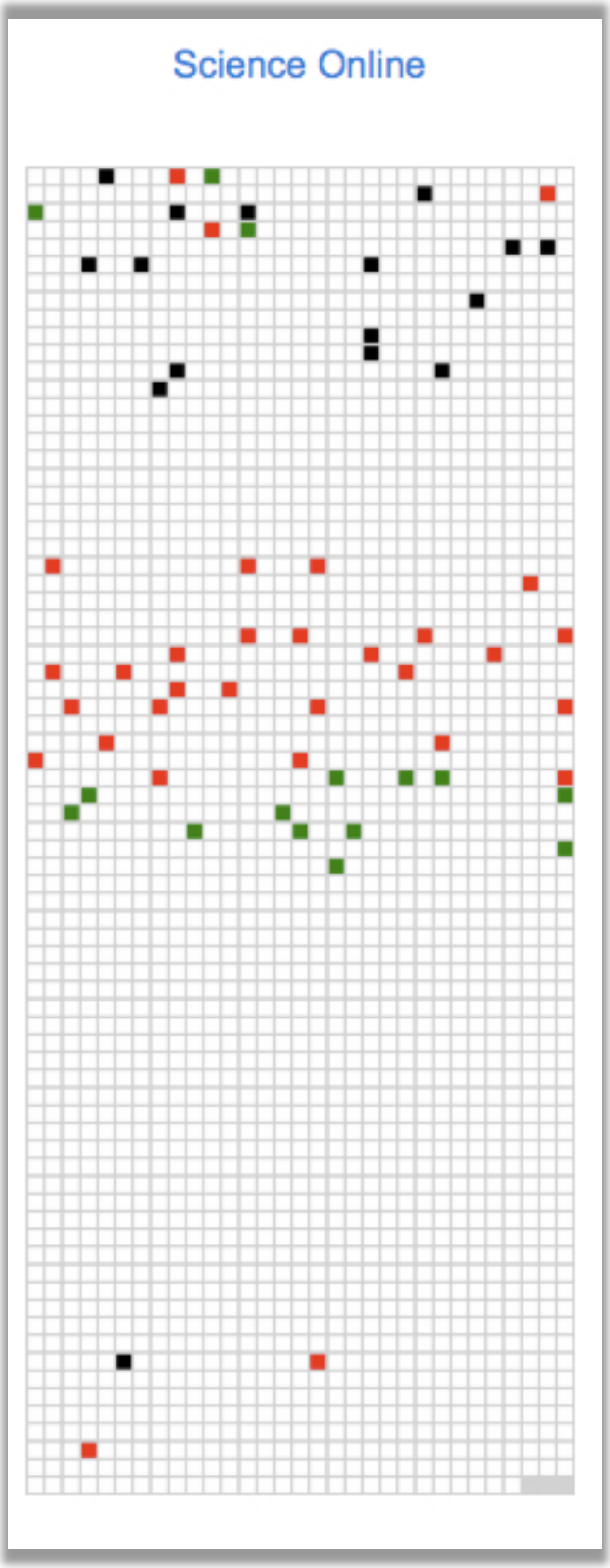
SV image for: wind wave solar

Entire web

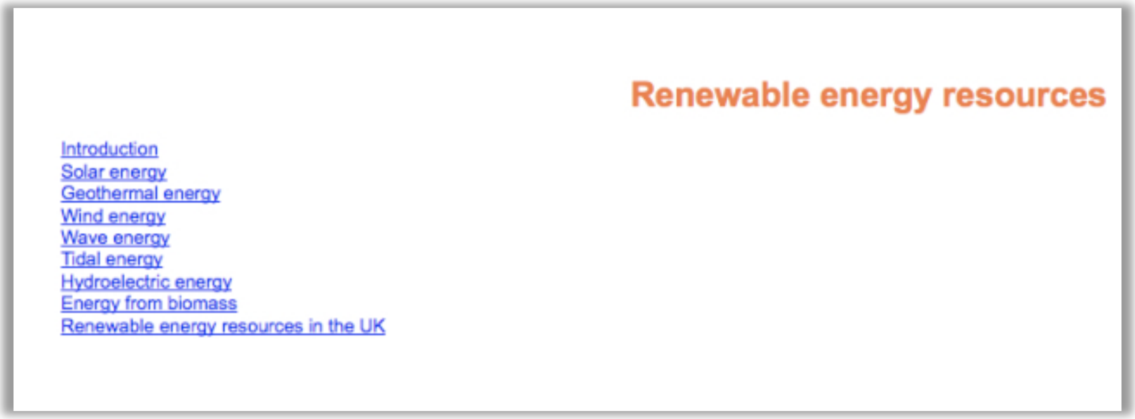
wind wave solar

There's a distinct pattern in the keyword distribution within the *Science Online* article at the right of the screenshot. All three terms are mentioned at the start of the article, but then there's a layer of mentions only of *solar*. That's followed by a short gap, after which there are mentions of *wind*, followed immediately by mentions of *wave*.

If we look at the whole of the article in this format, we see that the closing section partially mirrors the opening section, with mentions of *wind* and *solar* but not *wave*.



The obvious, and correct, conclusion is that the article is structured in well-defined sections. There's an introductory section which gives an overview of the topics covered below; then there are separate sections on each energy type. The gaps occur where the article describes energy types which are not *wind*, *wave* or *solar*. When we go to the article, we find the following header:



Renewable energy resources

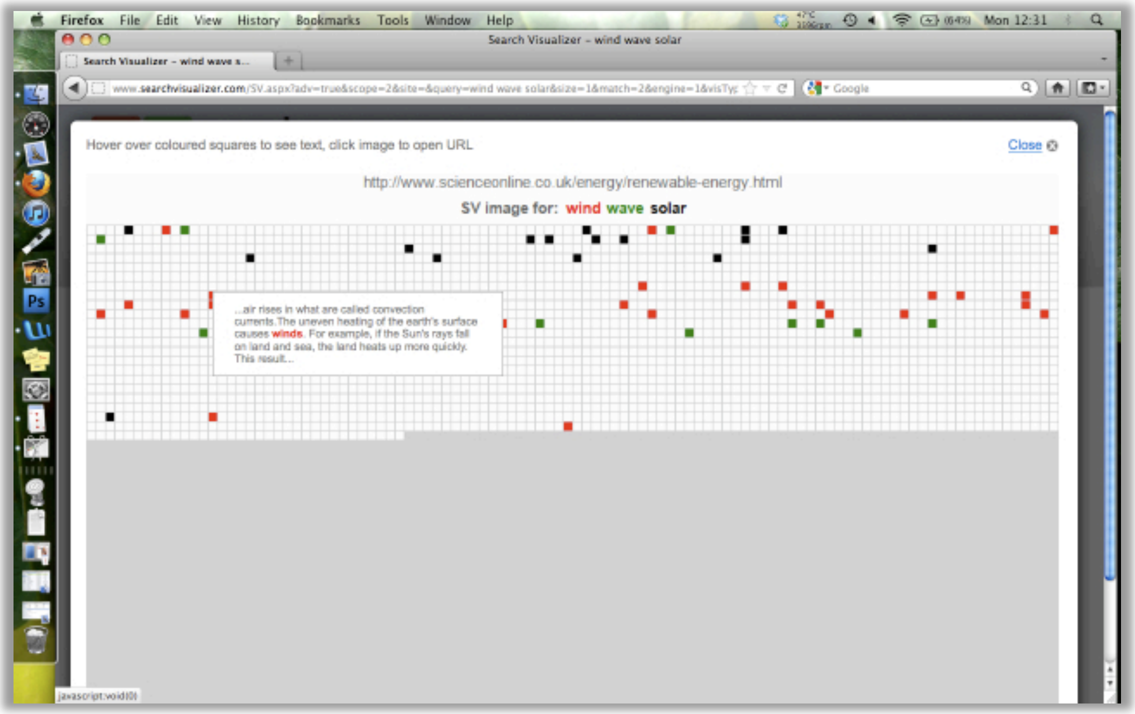
Introduction
Solar energy
Geothermal energy
Wind energy
Wave energy
Tidal energy
Hydroelectric energy
Energy from biomass
Renewable energy resources in the UK

(screenshot from: http://www.scienceonline.co.uk/energy/renewable-energy.html)

The first gap between layers of keywords is where there's a section on geothermal energy; the second, longer, gap occurs where there are sections on tidal energy, hydroelectric energy and energy from biomass.

In other words, the gaps are the places where the "not the usual suspects" topics are located.

When you've found a relevant-looking record, Search Visualiser lets you examine the visualization so you can decide whether or not you want to proceed to the article itself; if you hover over a keyword, Search Visualiser shows you the text surrounding that keyword. If you want to proceed to the article itself, you simply click on one of the white squares.


Firefox
File
Edit
View
History
Bookmarks
Tools
Window
Help
Mon 12:31
Search Visualizer – wind wave solar
Search Visualizer – wind wave s...
www.searchvisualizer.com/SV.aspx?adv=true&scope=2&site=&query=wind wave solar&size=1&match=2&engine=1&visTyp
Google
Close
javascript:void(0)

Hover over coloured squares to see text, click image to open URL

http://www.scienceonline.co.uk/energy/renewable-energy.html

SV image for: wind wave solar

...air rises in what are called convection currents.The uneven heating of the earth's surface causes winds. For example, if the Sun's rays fall on land and sea, the land heats up more quickly. This result...

One advantage of the Search Visualiser's format is that it can show very large documents on a single screen – for instance, it's possible to show an entire Shakespeare play in one screen.

This format has the added advantage of letting the user find relevant records even in languages that they don't speak; it's simply a case of translating the keywords into the target language, and then identifying relevant records using the same process as with the *wind wave solar* example above; relevant records can then be translated using online translation software. This can be very useful for researchers who want to know about specialist research in countries whose language they don't speak.

There's a substantial literature about the human ability to identify patterns, and about the difficulties of getting software to identify those same patterns. What we've done with Search Visualiser is to transform information into a format that plays to the strengths of the human being, particularly the strengths that the human hardly thinks about, such as knowing how a well-written document is structured.

There are other examples of how this can be used in the "Texts and Resources" section of the Search Visualiser site:

www.searchvisualiser.com