

Search Visualiser, pattern matching, and parallel processing, part 1

By Gordon Rugg

Background: This article unpacks some of the underlying concepts on which Search Visualiser is based.

This article was originally posted on March 8, 2012

A key feature of SV's design is the way it divides tasks between the computer and the human being. Computers are very good at what's known as *sequential processing* – that is, doing one step at a time. Humans aren't particularly good at sequential processing, but they are very good at *parallel processing*, where you do several things simultaneously. The concept of parallel processing goes beyond everyday multi-tasking, into things that are so familiar that we hardly ever think about them, but that make a huge difference for software design.

Here's an example which looks trivially easy. What are the yellow things in this photo?



(Image copyright Gordon Rugg, 2012)

To a human, the answer is easy: they're flowers, tulips to be precise. However, for a piece of software, that question would be just about impossible to answer. Identifying what the objects are in a photo requires a skill called *pattern matching*, which is a seriously difficult problem for present-day software. This particular form of pattern matching requires a massive amount of parallel processing by the visual system (which most software currently can't handle), and requires a lot of *knowledge about the real world* (which is also a serious problem for software). Humans, however, are extremely good at pattern matching, and parallel processing, and at real world knowledge, so it makes sense to design software to play to those human strengths.

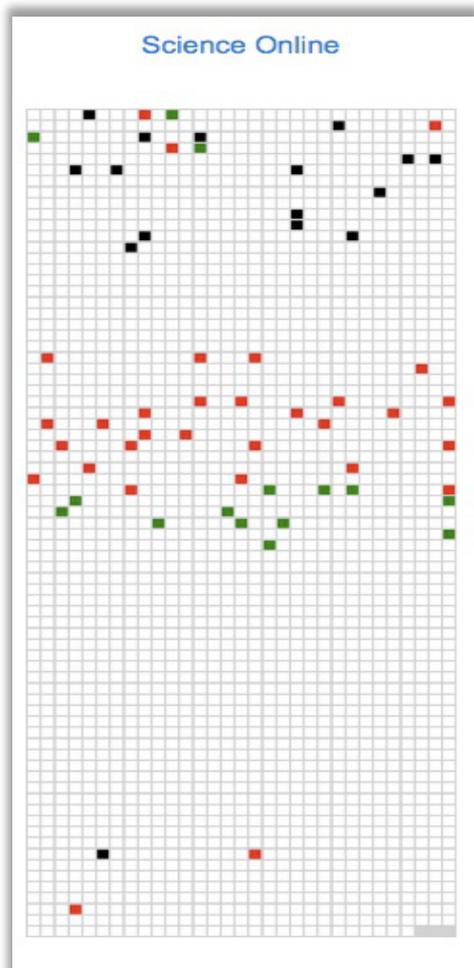
With SV, we've divided up the tasks of online search and text visualisation so that the computer does the parts that can be best handled by sequential processing, and so that the human does the parts that can be best handled by parallel processing and pattern matching.

The key place where that makes a significant difference is just after the search engine tells you that it's found several hundred thousand hits for your query. That first part of the search usually takes the computer less than a second.

Then things usually slow right down, because it's now your job to decide which of those several hundred thousand hits are relevant. If you try doing that by simply reading the records, you're using sequential processing, reading one word at a time. (The full story is more complicated, but that's a subject for a different post. In practice, though, the end result is the same: you're in effect reading one word at a time.) At a reading speed of a few hundred words per minute, it's going to take a long time to wade through even a fraction of those images via sequential processing.

With SV, however, the records are shown as images, so you can use parallel processing and pattern matching to identify relevant records much more swiftly. Here's an example. It's from a search about renewable energy sources, with the keywords *wind wave solar*.

What's going on in the image, in terms of pattern matching and parallel processing?



One feature which is immediately obvious to a human is the bands of colour. There's a black band near the top, then a gap, then a red band, then a green band immediately after the red one, then another gap.

For a human, that's very easy to identify via pattern matching. For software, it's a very difficult problem.

A subtler feature is what the distribution of colours means. To a human, with knowledge of how documents are usually structured in the real world, that's a fairly easy question to answer.

Each band of colour probably corresponds to a section of the document which is about a particular topic – for instance, the red band will probably be a section about wind power.

The beginning of the document includes all three keywords close to each other. What does that imply? Again, real world knowledge provides a likely answer. Documents often begin with a general overview of key topics, before going into detail about the individual topics one at a time. Also, documents often end with a conclusion which summarises the key findings. We'd therefore expect our keywords to occur together at the start and the end of the document, which is just what we find here. Interestingly, only *wind* and *solar* are mentioned in the end section, which implies that the document judges them to be more promising than wave power.

Taken together, that's a lot of information which a human can infer from a single image in a matter of seconds. A neat side-effect is that you can make the same type of inferences even if you don't speak the language of the original document, just by seeing the distribution of the keywords. Also, you can use the software without needing to go on a training course or read a manual, because it's showing the information in a format that you already understand because of your real world knowledge.

There's a lot more that you can do by using pattern matching and real world knowledge in combination with an SV display showing where your keywords occur. We've covered some of that in previous blog posts.

We'll be going deeper into parallel processing, pattern matching and real world knowledge in later posts. They have some far-reaching implications, that go well beyond online search and document visualisation.

Notes

The Search Visualiser is available for online use, free, at:
www.searchvisualiser.com