

Gardner, Mike (1999) A science-oriented search engine could solve web problems. *Nature*, 401 (111). p. 6749.

**Access from the University of Nottingham repository:**

[http://eprints.nottingham.ac.uk/1/1/www\\_science\\_search.pdf](http://eprints.nottingham.ac.uk/1/1/www_science_search.pdf)

**Copyright and reuse:**

The Nottingham ePrints service makes this work by researchers of the University of Nottingham available open access under the following conditions.

This article is made available under the University of Nottingham End User licence and may be reused according to the conditions of the licence. For more details see:

[http://eprints.nottingham.ac.uk/end\\_user\\_agreement.pdf](http://eprints.nottingham.ac.uk/end_user_agreement.pdf)

**A note on versions:**

The version presented here may differ from the published version or from the version of record. If you wish to cite this item you are advised to consult the publisher's version. Please see the repository url above for details on accessing the published version and note that access may require a subscription.

For more information, please contact [eprints@nottingham.ac.uk](mailto:eprints@nottingham.ac.uk)

School of Biomedical Sciences,  
Medical School,  
University of Nottingham,  
Nottingham. NG7 2RD

Sir,

Lawrence and Giles [1] eloquently define the current problems with the World-Wide Web, but could “Nature” provide the solution ?

The web has 800 million pages (and is still growing) but only 16% are indexed by the largest search engine. What is worse is that the newer search engines are biasing their index towards more ‘popular’ pages which may “delay or even prevent the widespread visibility of new high-quality information”. The diminishing returns in indexing the whole web suggest that a large proportion of the diverse array of scientific information available on the web, from research group home pages to preprints and sequences, may never be discovered.

Lawrence and Giles also point to a way forward: only 6% of web servers have scientific / educational content – a much more manageable amount to index. Metadata (information about information) is the key to better and more powerful searching, yet only 0.3% of servers use the Dublin Core [2] standard for metadata [1]. It seems clear that we need a science orientated search engine together with a scientific based set of metadata to help us trawl the oceans of information that *could* be available on our desktops.

I think that a hybrid indexing scheme, halfway between the formidable task of a human generated internet catalogue like Yahoo, and the huge computer generated indexes like Alta Vista is the best way forward for a scientific index. The submission of an institutions home page would, after manual verification (thus excluding inappropriate sites), lead to the computer generated indexing of the rest of the site.

The next step is a discussion by the scientific community to define suitable extensions to the Dublin Core metadata to describe the rich variety of scientific information available on the web. Nature’s debates [3] would seem to be the ideal forum for such a discussion.

Mike Gardner  
mike.gardner@nottingham.ac.uk

## References

- [1] Lawrence, S. and Giles, C.L. *Nature* **400**, 107-109 (1999).
- [2] Dublin Core. *The Dublin Core: A Simple Content Description Model for Electronic resources* (1999); <http://purl.oclc.org/dc/>
- [3] Nature. *Nature Debates* (1999); <http://helix.nature.com/debates>