This is a portion of the eBook at doi:10.7551/mitpress/8479.001.0001

Can Search Tame the Wild Web? Can Open Access Help?

From "Can search tame the wild web? Can open access help?" SPARC Open Access Newsletter, December 2, 2005.

http://dash.harvard.edu/handle/1/4727442

Which is growing faster, the size of the web or the power of search engines? No one knows, but it matters for all of us. If the power of search is growing faster, then it will eventually overtake the web's wildness and tame it—all of it. If the web is growing faster, then it will forever outpace our attempts to map, navigate, explore, and understand it.

The problem reminds me of the critical density of the universe. If there is enough matter in the universe, then gravitational attraction will eventually slow down and reverse cosmic expansion. The universe will end in a big crunch. If there isn't enough matter in the universe, the expansion will continue unimpeded indefinitely. The universe will end in heat death. When I was a student, I don't think we knew which outcome was more likely. But the recent discovery of accelerated expansion, perhaps due to dark energy, means that expansion is outpacing gravity.

We may not have a preference between the big crunch and heat death, especially if they're both billions of years in the future. But we should definitely root for the power of search to overtake the wild and rapid expansion of the web. If it does, then as readers we'll be able to find what we want and as authors we'll have a chance of being found by readers. But if search can't keep up with expansion, then as John Alan Paulos put it, the web will be the world's largest library, but its books will be scattered all over the floor. […]

Unfortunately it's difficult or impossible to make a direct comparison between the rate of web growth and the rate of search improvement. One reason is that web growth and search power are not forces of nature (limiting my analogy). Either or

Downloaded from http://direct.mit.edu/books/oa-edited-volume/chapter-pdf/2277836/9780262329552_cca.pdf by guest on 30 May 2024

This is a portion of the eBook at doi:10.7551/mitpress/8479.001.0001

both may suddenly speed up or slow down. We aren't surprised when they fluctuate because we know they are functions of a thousand variables that we do not yet fully understand.

There are many reasons why search may not keep pace with the expanding web. One is limited scope. What percentage of the web does a search engine index? A critical part of search scope is the dark web, a.k.a. the deep or invisible web stored in databases and closed to search engine crawlers. By one estimate the deep web is 500 times larger than the surface web.

Another is the adequacy of the relevance algorithm. How good is the search engine at guessing what you want and giving it a favored position in the search returns? Another is search spam. How good is the search engine at resisting attempts to manipulate its index and ranking algorithm? Another is lack of personalization. Even if the search index is complete and the relevance algorithm top-notch, different users may have different needs when searching on the same search term. Some will find what they want at the top of the relevance-sorted list but others won't, and it doesn't help them to know that what they seek is listed somewhere among the millions of other sites further down the list. (If you insist that search spam and lack of personalization are problems within the relevance algorithm, not separate problems beyond it, then I'll agree.)

The dark web is like dark energy, which accelerates expansion, not dark matter, which slows it down. By resisting search, the dark web helps the cause of entropy and hurts the cause of closure and organization.

Some of the literature in the dark web can still be searched by specialty search engines like Copernic, Deep Query Manager, iBoogie, and ProFusion. Moreover, even when dark web content is invisible to outside tools, it is usually hosted by databases that offer their own search functions. But even together these tools don't take us very far. Most of the dark web is still an unmapped region where content expands unnoticed by web-wide search engines.

Price barriers seclude content just as databases do and create a similar kind of web darkness. Some publishers of non-OA content invite major search engines like Google and Yahoo to index them, and I applaud that. And some search engines, like Scirus, already cover segments of the non-OA literature because they are created by the publishers. But most toll-access literature is still invisible to the general search engines.

In this sense, open access contributes to the mapping, taming, organization, and intelligibility of the web, while toll access resists the forces of closure.

Search is the gravity of the online universe that holds everything together, if anything does. As such, open access is one of its greatest allies. One thing we can do to defeat web entropy and help the cause of organization and discovery is to provide open

Downloaded from http://direct.mit.edu/books/oa-edited-volume/chapter-pdf/2277836/9780262329552_cca.pdf by guest on 30 May 2024

This is a portion of the eBook at doi:10.7551/mitpress/8479.001.0001

access to a continually growing proportion of research literature. The harder this literature is to index—and price barriers increase the difficulty—the harder it is for search to keep pace with the relentless expansion.

Postscript

I haven't seen other discussions of the analogy between cosmic expansion and web expansion, and the power of gravity and the power of search. But I can't believe I'm the first to notice it. If anyone can point out earlier discussions, I'd be glad to credit the pioneers.

Downloaded from http://direct.mit.edu/books/oa-edited-volume/chapter-pdf/2277836/9780262329552_cca.pdf by guest on 30 May 2024

This is a portion of the eBook at doi:10.7551/mitpress/8479.001.0001

Downloaded from http://direct.mit.edu/books/oa-edited-volume/chapter-pdf/2277836/9780262329552_cca.pdf by guest on 30 May 2024

This is a section of doi:10.7551/mitpress/8479.001.0001

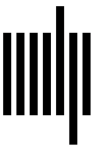
Knowledge Unbound

Selected Writings on Open Access, 2002–2011

By: Peter Suber

Citation:

Knowledge Unbound: Selected Writings on Open Access, 2002–2011
By: Peter Suber
DOI: 10.7551/mitpress/8479.001.0001
ISBN (electronic): 9780262329552
Publisher: The MIT Press
Published: 2016



The MIT Press

Downloaded from http://direct.mit.edu/books/oa-edited-volume/chapter-pdf/2277836/9780262329552_cca.pdf by guest on 30 May 2024

MIT Press Direct

© 2016 SPARC
This work as a collective, edited work is licensed to the public under a Creative Commons Attribution 4.0 International license:
http://creativecommons.org/licenses/by/4.0/

This work incorporates certain previously published materials. Copyright in some of those materials is owned by SPARC, and in others by Peter Suber. All were published under a CC-BY license.

This book was set in ITC Stone by Toppan Best-set Premedia Limited. Printed and bound in the United States of America.

Library of Congress Cataloging-in-Publication Data.

Names: Suber, Peter, author.
Title: Knowledge unbound : selected writings on open access, 2002-2011 / Peter Suber ; foreword by Robert Darnton.
Description: Cambridge, MA ; London, England : The MIT Press, [2015] | Selection of writings, mostly from the author's SPARC open access newsletter. | Includes index.
Identifiers: LCCN 2015038285 | ISBN 9780262029902 (hardcover : alk. paper) | ISBN 9780262528498 (pbk. : alk. paper)
Subjects: LCSH: Open access publishing. | Communication in learning and scholarship—Technological innovations.
Classification: LCC Z286.O63 S822 2015 | DDC 070.5/7973—dc23 LC record available at http://lccn.loc.gov/2015038285

10 9 8 7 6 5 4 3 2 1

Downloaded from http://direct.mit.edu/books/oa-edited-volume/chapter-pdf/2277836/9780262329552_cca.pdf by guest on 30 May 2024