ICHIM
PARIS 21-23 SEPT. 05
www.ichim.org

Digital Culture & Heritage
Patrimoine & Culture Numérique

Bibliothèque nationale de France, PARIS
Sept. 21st - 23rd, 2005
21 - 23 septembre 2005

FAUT-IL CENTRALISER LES DONNEES NUMERIQUES POUR AMELIORER LA RECHERCHE ?

Jerome Pesenti

Vivisimo Inc. USA

Published with the sponsorship of the French Ministry of Culture and Communication

Actes publiés avec le soutien de la Mission de la Recherche et de la Technologie du Ministère de la Culture et de la Communication, France

Interprétation simultanée du colloque et traduction des actes réalisées avec le soutien de l'Agence Intergouvernementale de la Francophonie

Abstract (EN)

Google, the new standard in information retrieval, wants to organize the world's information and make it universally accessible and useful, effectively centralizing all the information on its gigantic cluster of computers. In doing so, Google faces technical, ethical as well as usability challenges.

For example, by giving access to multiple types of contents (images, audios, videos, shopping, news, etc) it is faced with a critical relevancy problem: what should a search on "Britney Spears" return first? a web page, a video, an audio file, a picture? The current answer is to create vertical searches accessible through tabs or different front ends, but their very low usage demonstrates that they offer a very limited solution.

By taking content out of the hand of their owners, Google also encounters serious privacy, security and copyright issues. The recent backslashes against the numerisation project and Google scholar, the fears raised by Google Earth and the recent controversial CNET article showing how to find private information about Google's CEO just by using Google, are all symptoms of that problem.

These challenges are not limited to Google, any major search project - "search portal" - universal library, company wide access to corporate information, government portal, etc - is faced with the same issues.

The web is by essence decentralized, there is no central server, no central authority and everyone can be become their own publisher. Web search engines go somewhat against these trend, they try to gather the world's information into their giant indexes, acting as de facto central sources of information. But does it have to be so? We will explore how new standards and new technology can bring us closer to the ultimate goal of universal information access and organization without necessarily requiring the burden and issues of universal centralisation.

Keywords: search engine, federated search, web services, document clustering.

