Information Processing and Management 48 (2012) 1067–1077

Contents lists available at SciVerse ScienceDirect

Information Processing and Management

journal homepage: www.elsevier.com/locate/infoproman

# An agenda for green information retrieval research

Gobinda Chowdhury *

*Centre for Information and Knowledge Management, University of Technology, Sydney, Australia*

ARTICLE INFO

*Article history:*
Received 1 September 2011
Received in revised form 18 January 2012
Accepted 2 February 2012
Available online 7 March 2012

*Keywords:*
Green IR
Cloud computing
Information retrieval
Green IT
Environmental sustainability
Sustainable development

ABSTRACT

Nowadays we use information retrieval systems and services as part of our many day-to-day activities ranging from a web and database search to searching for various digital libraries, audio and video collections/services, and so on. However, IR systems and services make extensive use of ICT (information and communication technologies) and increasing use of ICT can significantly increase greenhouse gas (GHG, a term used to denote emission of harmful gases in the atmosphere) emissions. Sustainable development, and more importantly environmental sustainability, has become a major area of concern of various national and international bodies and as a result various initiatives and measures are being proposed for reducing the environmental impact of industries, businesses, governments and institutions. Research also shows that appropriate use of ICT can reduce the overall GHG emissions of a business, product or service. Green IT and cloud computing can play a key role in reducing the environmental impact of ICT. This paper proposes the concept of Green IR systems and services that can play a key role in reducing the overall environmental impact of various ICT-based services in education and research, business, government, etc., that are increasingly being reliant on access and use of digital information. However, to date there has not been any systematic research towards building Green IR systems and services. This paper points out the major challenges in building Green IR systems and services, and two different methods are proposed for estimating the energy consumption, and the corresponding GHG emissions, of an IR system or service. This paper also proposes the four key enablers of a Green IR viz. Standardize, Share, Reuse and Green behavior. Further research required to achieve these for building Green IR systems and services are also mentioned.



## 1. Introduction

Information retrieval (IR) which was once of interest and concern for a select few, especially those who were engaged in knowledge-intensive activities like education and research, has now become an integral part of our everyday activities. Every time we look for information on the web, in a digital library or in a database of some kind, we use an IR tool like a search engine. Although the use of information retrieval tools has proliferated over the past two decades since the advent and proliferation of the web and digital libraries, the IR research and industry existed for nearly five decades that have been engaged in building tools and techniques "to improve the process of finding information not only on the web, but also within a single computer ("desktop search") or set of computers ("enterprise search"), as well as within very large databases, such as libraries ("database search"). Further, IR techniques have been used to identify key links within, for example, legal records, genomics data, and spam" (Rowe, Wood, Link, & Simoni, 2010, p. 19).

* Tel.: +61 295142712.
*E-mail address:* Gobinda.chowdhury@uts.edu.au

0306-4573/$ - see front matter Crown Copyright © 2012 Published by Elsevier Ltd. All rights reserved.
doi:10.1016/j.ipm.2012.02.003

Nowadays we conduct billions of searches every month to find information on the web, library catalogs and databases, digital libraries, institutional repositories, e-books, e-journals, and so on. Behind every search there is an IR system that provides us access to the information that we seek. IR systems make extensive use of ICT (Information and Communication Technologies) at every stage of their creation and management to use, and such extensive use of ICT has an adverse environmental impact (Chowdhury, 2011). Creation as well as destruction of ICT equipments create harmful gases, called GHG (greenhouse gases, a term used to measure environmental impact of a product or service which is discussed later in the paper), and in order to run ICT equipments we need to use energy which also generates a substantial amount of GHG. To put it in a different way, as in our daily life we are becoming more and more reliant on IR systems for accessing and using digital information in a variety of forms, we are creating more environmental damages because of the increased use of ICT and energy. Consequently, it may be argued that we need to find a way to develop IR systems and services that are environmentally sustainable.

In this paper the term information retrieval is used in its broader context: it incorporates all the activities involved in the process of information access ranging from input creation, and processing to search, retrieval and use. In order to be sustainable, IR should be Green, and such a Green IR system or service should produce minimum GHG, in all its activities ranging from input creation to processing to information access and use. This paper aims to address the basic question of how can we build a Green IR that is sustainable, and in doing so it aims to build a Green IR research agenda. This paper first discusses the issue of climate change and environmental sustainability. Along this line it then proposes the concept of Green IR which should be economically, environmentally and socially sustainable. Although this paper primarily focuses on the environmental sustainability issue, it is argued that by building Green IR systems and services in accordance with the principles of Green IT/IS (information technology/ information systems) and cloud computing technologies, it is possible to achieve some degree of economic sustainability. Furthermore, the issue of social sustainability of Green IR is also discussed especially in relation to the concept of Green user behavior which is one of the conditions for building Green IT/IS. In order to build a case for a Green IR research, the paper looks at various related research issues in the areas of sustainable development with special reference to selected environmental sustainability, Green IT and cloud computing literature chosen from Scopus and ISI Web of Knowledge databases. It also looks at some reports on cloud computing such as those produced by the US National Institute of Standards and Technology, the Berkeley report on cloud computing, and the recent JISC (Joint Information Systems Committee in UK) initiatives on cloud computing, the *Hargreaves Review* of IP (intellectual property) laws in Britain, and some novel research approaches taken by SURFnet to promote online collaboration in the education and research sector in The Netherlands. It is proposed that the LCA (lifecycle analysis) approach which is used for estimating the GHG emissions of a product or service can be used to calculate the energy consumptions and consequently the GHG emissions of an IR system or service. However, given the complexity and resource-intensive nature of the LCA approach especially for measuring the energy consumptions of an IR system or service, an alternative and relatively simple approach, which has been taken in an UC Berkeley study for estimating the energy consumption of the Internet (Raghavan & Ma, 2011), may be a better choice.

## 2. Sustainability and climate change issues

As discussed earlier, environmental impact of a product or service is measured in terms of GHG emissions. There are many definitions of GHG some of which only talk about the emission of carbon dioxide ($CO_2$) but a broader definition covers emission of not only $CO_2$ but other harmful gases like nitrous oxide, ozone, hydrocarbon and chlorofluorocarbons, plus black carbon (Wiedmann & Minx, 2008). However, often GHG emission is expressed in metric tonnes (1000 kg) of $CO_2$ equivalent ($mTCO_2e$) which is an aggregated figure for emissions of all harmful gases but converted to $CO_2$ equivalent (IPCC, 2007a).

Sustainable development has remained an area of concern, and a major UN research and policy agenda for several years. The UN Conference on the Human Environment, authorized by the UN General Assembly in 1968, was held in Stockholm in 1972 where "Sustainability was a major theme, expressed as the idea that it was possible to combine economic growth with environmental protection" (Nolin, 2010). Although the issue of environmental sustainability has drawn much attention over the past few years, it is just one component of sustainable development, the other two are economic sustainability and social sustainability. They are interdependent because in order to achieve a sustainable development, we need to build systems and services that are not only environmentally sustainable but are also economically and socially sustainable. Different parameters for achieving economic and social of sustainability are listed in the site of United Nations Division for Sustainable Development (UNDSD; UN, 2009). However, although access to, and appropriate use of, information is a key to sustainable development, the importance of information in sustainable development has not been recognized or researched well either within the information community or within the wider community (Chowdhury, 2011; Nolin, 2010).

Off late environmental sustainability has become a major agenda item at every international, national and specific business/institutional level. The Intergovernmental Panel for Climate Change (IPCC) was created in 1988 by the United Nations Environment Programme (UNEP) and the World Meteorological Organization (WMO) as an international research and advisory body. IPCC is an intergovernmental body which currently has 194 member states. It is a scientific research and advisory body entrusted with reviewing and assessing scientific, technical and socio-economic information produced worldwide on the issue of climate change. Several studies and reports have been published under the IPCC banner over the past few years amongst which the *Climate Change 2007* report, also known as the *IPCC Fourth Assessment Report*, is the one that has been most widely known and discussed. It provides a detailed account of the GHG emission figures, discusses their future impact