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

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 on the world's climate, and points out the major areas of concern. It also proposes several guidelines for government, institutions and policymakers for combating the climate change issue. The *IPCC Fourth Assessment Report* has three parts:

– Part 1 "The Physical Science Basis" of the report discusses "the current understanding of the physical science of climate change that are judged to be most relevant to policymakers." (IPCC, 2007b). This report shows that climate change that has taken place in the recent past, and is taking place now, is mostly due to human activities.
– Part 2, "The Impacts, Adaptation and Vulnerability" report draws on over 290,000 data series to provide "a much broader set of evidence of observed impacts coming from the large number of case studies developed over recent years" (IPCC, 2007c). This report shows the impact of climate change on different areas and their vulnerability.
– Part 3 "Mitigation of Climate Change" of the report aims to answer five major questions:
  • "What can we do to reduce or avoid climate change?
  • What are the costs of these actions and how do they relate to the costs of inaction?
  • How much time is available to realize the drastic reductions needed to stabilize greenhouse gas concentrations in the atmosphere?
  • What are the policy actions that can overcome the barriers to implementation?
  • How can climate mitigation policy be aligned with sustainable development policies?" (IPCC, 2007d).

One of the landmark policy documents on climate change, that resulted from the IPCC assessment reports and subsequent UN and various government initiatives, is known as *The Kyoto Protocol*, which was adopted in Kyoto, Japan, on 11th December 1997 and entered into force on 16th February 2005. It has set binding targets for 37 industrialized countries and the European Community for reducing GHG emissions to an average of 5% against the 1990 levels over the five-year period of 2008–2012 (UNFCCC, 2011). Consequently, many countries now have developed new initiatives and strategies for reducing their emissions, and several targets have been set. As discussed below, appropriate use of ICT in various sectors has been considered as one of the ways of reducing GHG emissions of businesses and industries. However, systems and services that make extensive use of ICT also have to change in order to meet the target of reducing the GHG emissions of the overall business, industry and government. Research shows that the increasing use of ICT in its current form is not environmentally sustainable. It is estimated that ICT's own sector footprint currently stands at 2% of global emissions and it will almost double by 2020 (The Climate Group, 2008). The same report also suggests that appropriate use of Green IT can reduce "annual man-made global emissions by 15% by 2020 and deliver energy efficiency savings to global businesses of over EUR 500 billion" (The Climate Group, 2008).

Green IT/IS research shows how the carbon footprint of ICT can be reduced in different ways, for example, (1) by developing more sophisticated software and business systems that can help in the reduction of GHG emissions of businesses, (2) by developing new and improved technologies for manufacturing ICT equipments and infrastructure, and (3) by sharing computing and ICT infrastructure and thereby optimizing the use of computing and network resources and also reducing energy consumption of ICT (see for example, Jenkin, Webster, & McShane, 2011; Armbrust et al., 2009; Baliga, Ayre, Hinton, & Tucker, 2011).

However, to date there is hardly any study that estimates the carbon footprint of IR systems and services that, as discussed in the previous section, are increasingly becoming a common part of our daily life and activities. There may be two reasons for this. First it is difficult to estimate the carbon footprint of all the various forms of IR systems and services that we use every day in accessing a variety of information. Second, IR systems are often embedded within a broader service portfolio, for example, as a web search service or within a digital library, a social networking service, a library database service, etc. Therefore it may not be easy to calculate the carbon footprint of the IR service separately from the entire digital service.

This paper proposes that a lifecycle analysis (LCA; ISO-14040, 2006) approach may be taken to identify the various stages and components of an IR system or service, and thus to identify the corresponding factors and their contributions to the overall GHG emission figures of an IR system. This approach may be taken for conducting several case studies, for example, for open access and institutional repositories, for specific e-books and e-journals, etc. in order to determine the main contributors to GHG emissions and thus what measures need to be taken to reduce those emissions in order to build Green IR. The lifecycle analysis will also identify where and how the use of ICT can be optimized by using Green IT and cloud computing technologies which will eventually reduce the overall GHG emissions from an IR system and thus will help us build a Green IR. However, given the resource requirements and complexities of the underlying processes, an alternative and relatively simple process for estimation of energy consumption and GHG emissions of IR systems and services is also proposed.

3. Environmental impact of information: some basic facts and figures

Although so far very little systematic research has taken place to determine the environmental impact of information systems and services, some data can be gathered from related studies in order to develop a generic picture.

Although research and development on digital libraries and online information services/databases have proliferated over the past two decades, a large part of the information industry and services is still based on printed information resources that generate a substantial amount of GHG. Table 1 provides some estimates of the environmental impact of printed information

Table 1
CO_2 emissions from printed books.

Printed books market (annual figures)	CHG emissions in million tonnes	Equivalent to annual emissions from power stations[c]
UK: Volumes sold 235.6 million (Booktrade.info, 2010)	1.8[a]–2.4[b]	13.8–18.9
Germany: more than 1 billion produced (German Book... 2009)	7.46[a]–10.2[b]	58.7–80.3
USA: 4.15 billion books produced (Eco-Libris, 2007)	30.96[a]–42.3[b]	243.7–333
Australia books: 64.8 million (Nielsen Book, 2010)	0.48[a]–0.66[b]	3.8–5.2

[a] Based on CleanTech figures @ 7.46 kg/book (Ritch, 2009).
[b] Based on Babcock School of Management figures @10.2 kg/book (Ritch, 2009).
[c] Based on a typical power station in the UK (CARMA, 2007).

sources (for details see Chowdhury, 2010, 2011, 2012). By taking the emission figures from Table 1, it can be estimated that the GHG emissions only from book productions in the US accounts for about 0.6% of the GHG emissions of the entire country, and the annual GHG emissions only from book productions in the US and Germany alone is equivalent to about 0.1% of the annual GHG emissions of the entire world (Chowdhury, 2012). This clearly suggests that printed content generates a substantial amount of GHG and thus in order to build a Green information service we should go for dematerialization, i.e. replace printed content with their digital counterpart (Chowdhury, 2011). However, this will make more demand on the Internet and IR systems, and consequently on ICT which will increase energy consumptions and therefore will generate more GHG.

But how much energy is consumed by ICT and how much GHG does it generate? As stated earlier in this paper ICT accounts for 2% of global emissions. Estimating the energy consumption of the internet is a complex process because of the global dimension of the internet and the arrays of equipments and tools used to build, manage and access Internet and the various services. Acknowledging the complexity of the task and possible room for errors, a recent study at UC Berkeley (Raghavan & Ma, 2011) estimates that the Internet consumes between 170 and 307 GW (GigaWatt) of electricity which is equivalent to 1.1–1.9% of the total energy usage of humanity (estimated to be 16 TW (terra watt)). It is also estimated that 53% of the internet's total energy use comes from embodied power, i.e. the energy consumed in the lifecycle (from production to destruction) of various equipments (computers, cables, switches, modems, etc.) used for the creation, management and use of the Internet.

A study by the Australian Computer Society (2010) noted that in 2009 Australia's ICT users consumed 13.248 million KW h of electricity which generated 14.248 million tonnes equivalent of CO_2 emissions which is nearly 2.5% of Australia's total emissions (which is 539 million tonnes). The Berkeley study also (Raghavan & Ma, 2011) notes that desktops and laptops comprise roughly half of the Internet's total power consumption. This has an important consequence for Green IR research which is discussed later in this paper.

In 2010 Google's overall consumption of electricity was reported to be 2.26 million MW h (Albanesius, 2011). Since Google has operations and major servers in many countries and they have different types of energy sources with different GHG emission rates, it is difficult to accurately translate this energy consumption figure to its GHG emission equivalent. However, taking the GHG emissions of the power stations in the UK where 1785 power stations generate 370,000,000 MW h energy and produce 227,000,000 tonnes of CO_2 (CARMA, 2007), it can be estimated that Google's annual GHG emission is 1.386 million tonnes of CO_2. This is a conservative estimate because power stations in many countries generate more GHG per MW h of electricity (compared to those in the UK) and it depends on the source of power (coal, nuclear, wind, etc.). Nevertheless, this gives us an impression of the overall carbon footprint of Google operations. Off late, Google is moving towards using more environment-friendly power sources for its operations which will eventually reduce the carbon footprint of its operations.

But how can we relate the overall energy consumption of Google to its IR operations? It is estimated that approximately 34,000 Google searches are conducted per second and at this rate just over 88 billion Google searches are conducted per month and just over a trillion (1057 billion) searches per year. At this rate one Google search of generates nearly 1.4 g of CO_2. The actual figure will be slightly less than this because although searching is the main business activity of Google it has many other services and they too consume electricity. However, this is higher than the figures provided by Google sometime ago, as reported in the following BBC news:

> "US physicist Alex Wissner–Gross claims that a typical Google search on a desktop computer produces about 7 g CO_2. However, these figures were disputed by Google, who say a typical search produced only 0.2 g of carbon dioxide." (BBC News, 2009)

There is a difference between the two estimations shown in the news item: the first one (by the US physicist) provides an estimate of a typical IR session on Google while the second one (provided by Google) estimates the carbon footprint of a specific search. Accurate estimation of the energy consumption and energy costs of a specific IR session and thus the overall energy consumption and environmental impact of an IR system or service is a complex process because one should not only get the energy consumption figures of the service provider (like Google in this example), but it is more important, and equally difficult, to estimate the embodied energy cost, i.e. the cost of all the equipments and services used to build and manage a search service as well as the energy cost of all the equipments and ICT facilities that are used to access the service, e.g.

the servers, desktops, laptops, networks, routers, modems, etc., plus the real time socket energy costs for each of these equipments and facilities. Nevertheless, Green IT and cloud computing can play a big role in reducing the embodied energy or embodied costs.

4. Green IT and cloud computing

Green IT refers to the initiatives and programmes that directly or indirectly address environmental sustainability (Jenkin et al., 2011). The objective of Green IT is to reduce the overall environmental impact of ICT by adopting a number of measures ranging from taking environment-friendly approaches to the production and use of ICT equipment and facilities to optimizing the use of ICT equipments and network infrastructure in order to reduce the energy consumption at every stage.

Cloud computing is considered to be one of the most appropriate options to realize the goal of Green IT which is to reduce the economic and environmental costs of ICT by sharing of computing and network resources (see for example, Hayes, 2008; Vaquero, Rodero-Merino, Cáceres, & Lindner, 2009). The US National Institute of Standards and Technology defines cloud computing as follows:

> *Cloud computing is a model for enabling ubiquitous, convenient, on-demand network access to a shared pool of configurable computing resources (e.g. networks, servers, storage, applications, and services) that can be rapidly provisioned and released with minimal management effort or service provider interaction.* (Mell & Grance, 2011)

Cloud computing services are available at different levels ranging from the use of a shared computing infrastructure, known as the *Infrastructure as a Service* or *IaaS* in short, to sharing hardware and other computing resources for building applications, known as the *Platform as a Service* or *PaaS* in short, to sharing all the computing and network resources including the software, known as the *Software as Service* or *SaaS* in short (Chang, Tsai, Huang, & Huang, 2010; Mell & Grance, 2011; Open Cloud manifesto, 2009). Each kind of cloud service involves different levels of sharing of resources and may be appropriate for different types of computer and software/system solutions.

The economic and environmental benefits of cloud computing derive from the shared use of computing and network resources, and, depending on the level of service chosen, there may be different degrees of shared use of the network, hardware, software as well as specific applications and tools, etc. On the one hand the opportunity for the shared use of resources save the consumers from having to make significant initial investments for setting up ICT infrastructure, but on the other because of the sharing the overall cost of computing becomes significantly reduced. Some figures for savings in terms of shared infrastructure and reduced energy costs that can be obtained from cloud computing are provided in the Berkeley report on cloud computing (Armbrust et al., 2009). The environmental benefits of cloud computing may derive from:

- Reduced server energy consumption which is achieved by optimizing the use of computing resources, i.e. using full computing power only when required, and this can be achieved by techniques such as, sleep scheduling and virtualization of computing resources (Baliga et al., 2011; Liu, Zhao, Liu, & He, 2009);
- reduced network energy consumption which is achieved by optimum utilization of resources, e.g. by very high volume of traffic which will justify the energy consumption of the network (Baliga et al., 2011; Mell & Grance, 2011) and
- reduced energy consumption at the client/consumer end which can be achieved by doing most, if not all, of the software and processing activities at the Cloud data center/service rather than at the user end, and thereby enabling users to use what is known as a ‘‘thin client’’ – a computing facility with minimum processing capabilities and minimum energy consumption (Baliga et al., 2011; Cervone, 2010; Mell & Grance, 2011).

Thus by using the cloud computing technologies it is possible to reduce the embodied energy costs and also to achieve a considerable degree of economic and environmental sustainability for ICT and consequently the IR systems and services that make extensive use of ICT. That’s why use of the cloud computing technology and architecture is being considered in different countries for providing sustainable information services. In the UK, the Joint Information Systems Committee (JISC) is promoting the idea of using the cloud computing technology for providing data and information access services for the higher education sector in the country (JISC, 2011a, 2011b, 2011c). Some universities in the US are now taking initiatives in developing cloud-based systems for managing research data and information. For example, the Computation Institute at the University of Chicago and Argonne National Laboratory are is working to develop a system called *Globus Online* ‘‘to implement data management logic, both Amazon and local storage, campus credentials for authentication, and a set of UChicago and Argonne researchers and their laboratories (both small and large, and from a range of disciplines) to evaluate effectiveness’’ (Foster, 2011). In The Netherlands, SURFnet, the national body for promoting ICT in education and research, is also taking several measures for promoting the use of cloud computing for higher education and research (SURFnet, 2011).

5. Cloud computing and IR research

A simple search conducted on 30.8.2011 on ‘‘cloud computing’’ and ‘‘information retrieval’’ returned 52 hits from Scopus and 69 hits from the ISI Web of Knowledge database. It was noted that this area of research is very new because the earliest result returned from Scopus was from 2008. The overall distribution of results from Scopus was: 15 papers published in 2011 (up to August 2011), 30 papers published in 2010, six papers in 2009 and only one in 2008. Again not all these papers were on cloud computing and information retrieval in strict sense of the terms, but they covered different issues of cloud computing and different issues of information management, information retrieval, business models for information and resource sharing etc. However, some research focused specifically on the information retrieval and access issues in the cloud computing environment. For example,

- Shih, Tseng, and Yang (2011) discuss the issue of retrieval of learning resources in the context of cloud computing,
- Bales, Sohn, and Setlur (2011) discuss information access behavior using mobile and other computing devices, especially smartphones and computers,
- Basirat and Khan (2010) propose that data storage and retrieval on the cloud can be performed using a distributed pattern recognition approach,
- Teregowda, Urgaonkar, and Giles (2010) propose that cloud implementation can improve the performance of CiteSeer, a service that provides citation indexing, full text indexing, and extensive document metadata from documents crawled from the web across computer and information sciences and related fields,
- Yang, Kamata, and Ahrary (2009) propose an open source cloud computing enabled content based image retrieval system called NIR, and
- Langer (2011) discusses how cloud computing can help researchers in the storage and retrieval of large medical image databases.

These show that IR researchers are increasingly finding cloud computing as a new opportunity for extending their research. However, as discussed earlier in the paper, there is lack of research on the environmental impact of IR systems and services in general, and more specifically how Green IR systems and services can be developed to promote the idea of environmental sustainability in the context of access to, and use of, digital information in different forms and in different application environments – education and research, business, government, etc. The main challenge comes from the lack of an established methodology for calculating the GHG emission figures for an IR system or service which is often embedded within a bigger product or service. The Life Cycle Analysis (LCA) approach which is used to estimate the energy consumptions and environmental impact of various products and services may be useful for developing a methodology for calculating the GHG emissions from an IR system as program or tool, and also from an information retrieval service that includes all the activities from processing of input information, i.e. digital content, to search, access and use of information.

6. Measuring the environmental impact of IR: the LCA approach

Although Green IT and cloud computing research is progressing, till date there is no authoritative research data on the environmental impact of IR systems and services. This is partly because of the complexities involved in the process of estimation, and therefore developing a Green IR research agenda and pursuing it is absolutely essential.

GHG emissions of a product or service is calculated by using what is known as the Life Cycle Assessment (LCA) (Finnveden et al., 2009; ISO-14040, 2006). This method takes into account the energy inputs and emission outputs throughout the production chain from exploration and extraction of raw materials to different stages of processing, manufacturing, storage, transportation, use and disposal. LCA is accredited by the ISO 14000 series standards that "reflects international consensus on good environmental and business practices that can be applied by organizations all over the world in their specific context" (ISO, 2009).

Although the software development lifecycle can be modified to achieve environmental sustainability, this is a relatively new area of research and there are not many research literature that specifically discuss how the LCA approach can be used to assess the GHG emissions from a software lifecycle. The International Workshop on Software Research and Climate Change (WSRCC) is a recently introduced series of workshops that addresses the environmental sustainability of software engineering (WSRCC, 2011). Galster (2010) notes that the LCA approach may be used to determine the environmental sustainability, or the carbon footprint (which is often used as an alternative term to denote GHG emissions), of:

- a particular software product,
- a software development process for an organization or a suite of consumer activities, or
- the whole software product life-cycle (development, usage, maintenance, etc.).

Each stage in the software development lifecycle should be analysed to determine how much energy is used, and the corresponding GHG emissions for production and use of every physical item that is used in the process, plus the energy cost for all the activities in the design, development, testing and implementation of the software including the energy consumption for the office space used, travels, if any, future maintenance including debugging, upgrading, etc.

Before conducting a lifecycle analysis it is important to define the scope of a product or service being studied. At one level, an IR system may be considered as a software or a search engine that allows users to search and retrieve digital information. An IR service as part of an information search service or a database such as an open archive, an institutional repository or a commercial database of content has the same goal as an IR system or a search engine, i.e. to provide access to information, but it has a broader scope in terms of LCA because it has to have a collection of digital content and it has a target user community that accesses and uses the digital information. Therefore the lifecycle of an IR as a tool and IR as a service is different and each will have different contributors for GHG emissions throughout its lifecycle. One may also argue that a search service may provide access to digital information but it may not have a collection of digital content, as in case of search engines. So, for a search engine the process starts from identification and indexing of digital content as opposed to content creation or harvesting, storage, processing, etc.

Thus for the purpose of LCA, it is important to consider three different classes of IR: (1) IR as a software or tool which is a product, (2) IR as a search service or a search engine that uses a product to provide search and retrieval facilities, and (3) IR as an integral part of an information service, for example a digital library, that uses a search engine that comprises an IR tool or software in order to facilitate access to, and use of, one or more collections of digital content for a designated user community. This classification is not always quite distinct and mutually exclusive, because IR as a software or a search service while playing part of an information access service like an online database, an open archive or an institutional repository, etc., is always evaluated and modified/improved as part of a usability testing cycle. Similarly there is a significant overlap in the scope of specific features of a search engine and a database search service, but for conducting the LCA to estimate GHG emissions it is useful to have this classification because for each class there may be several different factors contributing to the GHG emission and these need to be identified and appropriate measures need to be taken to reduce the GHG emissions from each component which will then lead to the development of Green IR.

7. Measuring the environmental impact of IR: an alternative approach

The LCA approach is quite resource intensive and depending on the level of analysis the process of estimation of energy consumption of an IR system or service can be quite complex. An alternative, and relatively less complex and less expensive, approach similar to that adopted for estimating the energy consumption of the Internet in a recent study at UC Berkeley (Raghavan & Ma, 2011), may be taken for estimating the energy consumption, and consequently the GHG emission figures, of an IR system or service. According to this approach, the energy cost of IR should be based on the estimation of two types of energy: (1) the embodied energy or emergy (Raghavan & Ma, 2011) of the different the devices – servers, desktops, laptops, networks, routers, modems, etc., used in building IR systems and services, such as databases or digital libraries and search engines, as well as the user devices such as desktops, laptops, ipads, ebook readers, and mobile devices that are used to access information, and (2) the socket energy, i.e. the energy consumed by various devices for the creation, storage and maintenance of the databases and search engines as well as the access and use of digital information. For example, for estimating the energy consumption of an IR system or service, it may be necessary to estimate:

- the various types of computing equipments and facilities used for development and management of the IR system/service, and energy required to manufacture, and end-of-life disposal, of those equipments and facilities
- the socket energy consumption of various equipments and facilities for creation and maintenance of the IR system or service
- energy consumptions for business, travel, office equipments and overheads in terms of office facilities, etc., used for development of the content vis-à-vis the IR system/service
- the types, number and duration of usage of different user devices for accessing the IR service and the corresponding energy requirements for manufacture, maintenance and disposal of those devices
- the socket energy consumption of various user devices.

Together these figures will give an estimation of the energy and environmental cost of an IR system or service. It is quite evident that the embodied energy cost per use/user will decrease as the number of access increases. However, the energy cost at the user end will increase as the number of user increases. This trade-off can be balanced by taking a number of measures, for example by optimization of the usage of computing and network resources by using the cloud computing model, by reducing the energy consumption at the client end by developing IR services that can be used by thin clients, and so on.

As stated earlier in this paper, the UC Berkeley study noted that embodied energy accounted for about 53% of the total energy consumption of the Internet, and also that desktops and laptops account for about half of the total energy consumption of the Internet. This may very well be the case for IR systems and services and thus embodied energy and especially desktops and laptops may be the major contributors to the overall energy consumption. While cloud computing may reduce the embodied energy consumption for IR systems and services, user behavior will play a major role because the socket energy consumption will depend on the duration of a search session and the type of computing facilities used, the source of energy used, and the overall search behavior of the user, for example whether the user reads the content and interacts with it online, or whether this is done in an offline mode, through printed content, and so on.

8. So, how can we develop a Green IR?

Whichever approach is taken the overall energy consumption, and the thus the GHG emission figures, of an IR system or service can be reduced by: (1) making IR systems more efficient so that the overall time for generation and processing of digital content by the IR system as well as the retrieval time is reduced, (2) making it possible for the user to access information using a thin client that has low energy consumption, and (3) reducing the socket energy consumption of the end-user devices. It should be noted that for the first option, there is a trade-off between the retrieval time and simultaneous use of multiple servers. For example, Google uses multiple servers simultaneously in order to reduce the retrieval time. While it has a significant impact on the efficiency of an IR service, retrieval time (measured in terms of the fraction of a second) may not be the specific requirement for every IR system. The second option is more important, and indeed cloud computing technologies are moving towards the use of thin clients at the user end so that the overall energy consumption at the user end can be reduced significantly. The third option is related to the second option, but creation and processing of digital content by IR systems can pay a major role here. For example, the time for access and use can be reduced by using granular content. This of course depends on the trends in the industry – both the content and the IT industry.

Furthermore, some generic measures may be taken to facilitate the development of Green IR. These may be grouped under four major headings and may be called the four key enablers of Green IR. These are Standardize, Share, Re-use and Green behavior (SSRGb):

1. *Standardize*: As noted earlier in this paper, the benefits of cloud computing technologies can be realized through more standardization in terms of content creation as well as content organization and processing. This has also been a specific recommendation of the recently released Hargreaves (2011) Review which is a government-sponsored study of the current state of IP (intellectual property) laws in Britain that stand in the way of innovation and progress in digital economy. Hargreaves (2011) makes some specific recommendations for reform of IP laws in Britain to make them more supportive of innovation and progress in knowledge-intensive activities. Two major recommendations that will have profound impact on IR are:
   a. Establishment of Digital Copyright Exchange to facilitate dynamic access to, and use of, digital content – commercial/copyright as well as non-commercial/free content, and
   b. Standardization in processes and practices: "By developing an open, standardized approach to data it will be possible to attach copyright conditions and rights information directly to digital content in a uniform machine readable fashion (so called meta data)" (p. 33)

These recommendations, once implemented, along with the potential of cloud computing for remote access to data and content through the PaaS and SaaS layers of the cloud architecture, will greatly facilitate IR activities. It is anticipated that by moving digital content and data to the cloud will significantly reduce GHG emissions that are currently produced for separate local data centers hosting data and content. However, the real benefits of this remote access facility can be realized if and only when standard metadata formats and tools are used to create and process/index digital data and content in order to make them interoperable for developing IR applications across various genres of content and data. Standardization of content creation using open standards will greatly facilitate IR developments ranging from content and data identification to indexing, and building several context and domain-specific applications using techniques of data/text mining, visualization, etc.

2. *Share*: As discussed earlier in the paper, the main ethos of cloud computing is sharing of resources. This is also the spirit of the Hargreaves Review that proposes several IP reforms and establishment of the Digital Copyright Exchange in order to facilitate sharing of content and knowledge for research and innovation. The IR community has already developed excellent examples of sharing resources and expertise through TREC initiatives for well over a decade. Furthermore, there have been some growing interests in crowd sourcing IR (a special track in TREC) and all these are excellent examples of sharing of IR resources and expertise. It is anticipated that by increasing the sharing of content, tools and expertise the IR community will be able to reduce a significant amount of GHG emission that is otherwise produced through isolated research efforts on data and content held on local servers. Different models of collaboration and sharing, especially using open standards and technologies, are now being developed and tested. One good example is the SURFconext (2011) collaboration infrastructure developed for collaboration in the higher education and research sector in The Netherlands. SURFconext, which is at its beta stage development, provides a good example of recent developments in open source collaboration infrastructure that can greatly facilitate collaboration and sharing and specific IR application development activities by integrating institutional resources with commercial cloud services. SURFconext will have applications from different developers and it will allow researchers, educators and students select the tools that best fit their online collaboration needs provided on its website (SURFconext, 2011).
3. *Re-use*: IR tools and content re-use has remained a major area of research within the IR and digital library research community for a long time. The TREC research community has for a long time re-used content and tools for innovative IR research. The cloud computing infrastructure will greatly facilitate this since various research data sets, tagged and indexed content, various indexes and tools etc., can be stored on the cloud for use in various research and application development activities. One of the major recommendations of the Hargreaves Review is:

   "Government should deliver copyright exceptions at national level to realize all the opportunities within the EU framework, including format shifting, parody, noncommercial research, and library archiving. The UK should also promote at EU level an exception to support text and data analytics." (Hargreaves, 2011, p. 8)

Such changes in the IP laws in favor of research and sharing of data and content will promote Green IR research because researchers do not need to spend computing and energy resources on something that has already been built, e.g. software applications, tagged and indexed collections, transaction log data, various data analytics, etc. and these can be re-used for further research and innovation.

4. *Green behavior*: Human information behavior has remained a major area of research within the information community and for many this is an integral part of IR research because the success or failure of any IR system or service depends on how it meets the user information needs which of course depends on a number of user behavior and characteristics, etc. Green user behavior has been recognized as an important component of Green IT/Green IS (Jenkin et al., 2011), and several projects have looked at the user behavior with regard to energy consumptions and user perceptions of their environmental impact (see for example a list of projects at JISC, 2011c). Green user behavior in the context of climate change may mean several things that include a range of behavior changes with regard to energy usage, business practices, lifestyle, and so on. This also relates to the social sustainability which, as discussed before, is an important enabler for environmental sustainability. At the very basic level, Green user behavior in the context of Green ICT may mean changes in the user behavior in relation to reducing printing and photocopying activities that have a significant adverse effect on the environment. It is estimated that printing and copying at UK universities account for 10–16% of ICT-related electricity consumption, and the total printing and copying costs in larger universities can go well over £1 m per year (James & Hopkinson, 2009). Proper user education supported by appropriate measures for easy access to information through the cloud can lead to the development of Green user behavior which is a key enabler for Green ICT. It is expected that copying and printing will decrease with the increasing sophistication and use of ipads, ebook readers, and various other user devices, as well as specific measures taken by IR systems and services where users can store, use and share content of their choice, for example, content that they have accessed before, where and when they want.

In addition to reducing printing and copying, information users need to change their overall business practices – education and learning activities, research and innovation, business and administration, etc. in such a way that digital content and data is accessed and used remotely using low end computing facilities. Research shows that significant reductions in terms of GHG emissions can be achieved only when users access and use data from the cloud using low energy computing devices (Baliga et al. 2011). However, more research and user education is needed to promote the idea of Green information user behavior, and thus achieve the social sustainability of IR systems and services.

9. Conclusion

IR has become an integral part of many, if not most, of our daily activities where we generate, access and use digital information in a number of ways for a number of our day-to-day activities. Given the new developments and promises of cloud computing, such as the recent JISC cloud computing initiatives in Britain discussed in this paper, and related developments such as the recent initiatives prompted by the Hargreaves Review will no doubt open new opportunities for new and more rigorous IR research. It is expected that these developments will encourage IR researchers to undertake research and develop new domain and context-specific IR tools and applications, including new data/text mining and visualization applications to facilitate access to integrated data and content from heterogeneous platforms and services. However, this increased research, and more frequent use of IR should be coupled with environmental research in order to develop sustainable IR systems and services.

Green IR that produces minimum GHG emissions in facilitating access to, and use of, digital information can significantly help in achieving sustainable development. Although a considerable amount of research has been undertaken in the recent past on Green IT, Green IS and cloud computing, to date hardly any research has been done on Green IR per se. In order to develop Green IR systems and services we first need to know which components of an IR system and service produces how much GHG and then we need to find appropriate ways of reducing those emissions. However, this calls for detailed research and analysis because IR systems vary quite significantly and they often form an integral part of a digital library or an information search service.

Concerted efforts from the IR research community are needed to promote Green IR as a major research agenda for building sustainable digital information and communication services which are increasingly becoming an integral part of our daily life and activities. Green IR research will help us find better ways to Standardize, Share and Re-use digital content as well as tools and technologies related to IR systems and services leading to environmental sustainability. It will also promote Green user behavior which will lead to social sustainability of Green IR systems and services. Appropriately designed digital information systems and services can play a key role in sustainable development, and appropriate Green IR research can play a big role in that.

References

Albanesius, C. (2011). How much electricity does Google consume each year? *PCMag.com*. <http://www.pcmag.com/article2/0,2817,2392654,00.asp> 11.09.11.

Armbrust, M., Fox, A., Friffith, R., Joseph, A.D., Katz, R.H., Konwinski, A., Lee, G. Patterson, D.A., Rabkin, A., Stoica, I., Zaharia, M. (2009). *Above the clouds: A Berkeley view of cloud computing*. Electrical Engineering and Computer Sciences, University of California Berkeley, Technical report No. UCB/EECS-2009-28. February 10, 2009. <http://www.eecs.berkeley.edu/Pubs/TechRpts/2009/EECS-2009-28.html> Retrieved 08.08.11.

Australian Computer Society (2010). Carbon and computers in Australia: The energy consumption and carbon footprint of IICT usage in australia in 2010. *A report for the Australian computer society by connection research*. Australian Computer Society. <http://www.acs.org.au/attachments/ICFACSV4100412.pdf> Retrieved 29.12.11.

Bales, E., Sohn, T., Setlur, V. (2011). Planning, apps, and the high-end smartphone: Exploring the landscape of modern cross-device reaccess. In *9th International Conference on Pervasive Computing, Pervasive 2011*. San Francisco, CA. 12 June 2011 through 15 June 2011; Code 85131. Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics) LNCS (Vol. 6696, pp. 1–18). Springer-Verlag.

Baliga, J., Ayre, R.W.A., Hinton, K., Tucker, R.S. (2011). Green cloud computing: Balancing energy in processing, storage, and transport. In *Proceedings of the IEEE* (Vol. 99). <http://ieeexplore.ieee.org/stamp/stamp.jsp?arnumber=05559320> Retrieved 08.08.11.

Basirat, A.H., Khan, A.I. (2010). Evolution of information retrieval in cloud computing by redesigning data management architecture from a scalable associative computing perspective. In *17th International Conference on Neural Information Processing, ICONIP 2010*. Sydney, NSW. 22 November 2010 through 25 November 2010. Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics). LNCS Issue PART 2 (Vol. 6444, pp. 275–282).

BBC News (2009). ''Carbon cost'' of Google revealed. <http://news.bbc.co.uk/2/hi/7823387.stm> 12.01.09.

CARMA: Carbon monitoring for action (2007). <http://carma.org/region/detail/201> Retrieved 12.01.12.

Cervone, H. F. (2010). An overview of virtual and cloud computing. *OCLC Systems and Services, 26*(3), 162–165.

Chang, B.R., Tsai, H.F., Huang, C.-F., Huang, H.-C. (2010). Private small-cloud computing in connection with Linux thin client. In *1st International conference on pervasive computing, signal processing and applications, PCSPA 2010* (pp. 82–87).

Chowdhury, G. (2010). Carbon footprint of the knowledge sector: what's the future? *Journal of Documentation, 66*(6), 934–946.

Chowdhury, G. (2011). Building environmentally sustainable information services: A green IS research agenda. *Journal of the American Society for Information Science and Technology*. doi:10.1002/asi.21703.

Chowdhury, G. (2012). How digital information services can reduce greenhouse gas emissions. *Online Information Review, 36*(4).

Finnveden, G., Hauschild, M., Ekvall, T., Guinée, J., Heijungs, R., Hellweg, S., et al (2009). Recent developments in Life Cycle Assessment. *Journal of Environmental Management, 91*(1), 1–21.

Foster, I. (2011). Research data lifecycle management as a service. <http://www.columbia.edu/~rb2568/rdlm/Foster_UChicago_RDLM2011.pdf> Retrieved 08.08.11.

Galster, M. (2010). Life-cycle assessment in software engineering. position paper <http://www.cs.toronto.edu/wsrcc/WSRCC2/papers/wsrcc2010-Galster.pdf>.

Hargreaves, I. (2011). Digital opportunity: A review of intellectual property and growth. An independent report. <http://www.ipo.gov.uk/ipreview-finalreport.pdf> Retrieved 08.08.11.

Hayes, B. (2008). Cloud computing. *Communications of the ACM, 51*(7), 9–11.

IPCC (2007a). IPCC climate change 2007: Synthesis report. <http://www.ipcc.ch/pdf/assessment-report/ar4/syr/ar4_syr.pdf> Retrieved 08.08.11.

IPCC (2007b). Climate change 2007: Working group I: The physical science basis <http://www.ipcc.ch/publications_and_data/ar4/wg1/en/frontmatterspreface.html>.

IPCC (2007c). Climate change 2007: Working group II: Impacts, adaptation and vulnerability. <http://www.ipcc.ch/publications_and_data/ar4/wg2/en/frontmattersforeword.html> Retrieved 08.08.11.

IPCC (2007d). Climate change 2007: Working group III: Mitigation of climate change. <http://www.ipcc.ch/publications_and_data/ar4/wg3/en/frontmatterspreface.html> Retrieved 08.08.11.

ISO (2009). Environmental management: the ISO 14000 family of international standards. <http://www.iso.org/iso/theiso14000family_2009.pdf> Retrieved 08.08.11.

ISO-14040 (2006). Environmental management – Life cycle assessment – Principles and framework. In *International Organization for Standardization*. Geneva: International Organization for Standardization.

James, P., Hopkinson, L. (2009). Green ICT: Managing sustainable ICT in education and research. <http://www.jisc.ac.uk/publications/programmerelated/2009/sustainableictfinalreport.aspx> Retrieved 08.08.11.

Jenkin, T. A., Webster, J., & McShane, L. (2011). An agenda for green information technology and systems research. *Information and Organization, 21*(1), 1–24.

JISC (2011a). Cloud computing increasingly attractive to universities, says JISC 16.5.2011 <http://www.jisc.ac.uk/news/stories/2011/05/cloud.aspx> Retrieved 08.08.11.

JISC (2011b). Cloud services for education and research – projects and partners announced, 23.6.2011 <http://www.jisc.ac.uk/news/stories/2011/06/cloudservices.aspx> Retrieved 08.08.11.

JISC (2011c). Green ICT: managing environmentally sustainable ICT in education and research. Energy dashboard and user behaviour, June 30, 2011 <http://greenict.jiscinvolve.org/wp/2011/06/30/energy-dashboards-and-user-behaviour/> Retrieved 08.08.11.

Langer, S. G. (2011). Challenges for data storage in medical imaging research. *Journal of Digital Imaging, 24*(2), 203–207.

Liu, J., Zhao, F., Liu, X., He, W. (2009). Challenges towards elastic power management in internet data centers. In *Proceedings of the IEEE international conference on distributed computer systems workshops* (pp. 65–72). Los Alamitos, CA.

Mell, P., Grance, T. (2011). The NIST definition of cloud computing: Recommendations of the National Institute of Standards and Technology NIST Special Publication 800-145 <http://csrc.nist.gov/publications/drafts/800-145/Draft-SP-800-145_cloud-definition.pdf> Retrieved 12.08.11.

Nolin, J. (2010). Sustainable information and information science. *Information Research, 15*(2), <http://informationr.net/ir/15-2/paper431.html> Retrieved 08.08.11.

Open cloud manifesto (2009) <http://www.opencloudmanifesto.org/opencloudmanifesto1.htm> Retrieved 08.08.11.

Raghavan, B., Ma, J. (2011). The energy and emergy of the internet. In *Proceedings of the ACM workshop on hot topics in networks (Hotnets)*. Cambridge, MA, November 2011. <www.cs.berkeley.edu/~jtma/papers/emergy-hotnets2011.pdf> Retrieved 12.01.12.

Ritch, E. (2009). *The environmental impact of Amazon's Kindle: Executive brief*. San Francisco, Cleantech Group.

Rowe, B. R., Wood, D. W., Link, A. N., Simoni, D. A. (2010). Economic impact assessment of NIST's Text REtrieval Conference (TREC) program. Final Report. RTI Project Number 0211875. <http://trec.nist.gov/pubs/2010.economic.impact.pdf> Retrieved 12.01.12.

Shih, W.-C., Tseng, S.-S., Yang, C.-T. (2011). Due time setting for peer-to-peer retrieval of teaching material in cloud computing environments. In *2011 International conference on information science and applications, ICISA 2011; Jeju Island; 26 April 2011 through 29 April 2011; 2011 international conference on information science and applications, ICISA 2011*, Article number 5772397 ISBN: 978-142449224-4, doi:10.1109/ICISA.2011.5772397.

SURFCONEXT (2011). <http://www.surfnet.nl/en/Thema/coin/whatis/Pages/default.aspx> Retrieved 08.08.11.

SURFnet (2011). Cloud computing. <http://www.surfnetkennisnetproject.nl/innovatie/cloudcomputing> Retrieved 31.08.11.

Teregowda, P., Urgaonkar, B., Giles, C. L. (2010). Cloud computing: A digital libraries perspective. In *3rd IEEE international conference on cloud computing, CLOUD 2010*. Miami, FL; 5 July 2010 through 10 July 2010. Proceedings – 2010 IEEE 3rd international conference on cloud computing, CLOUD 2010, 2010 (pp. 115–122) Article number 5558004.

The Climate Group (2008). SMART2020: Enabling the low carbon economy in the information age. <http://www.theclimategroup.org/publications/2008/6/19/smart2020-enabling-the-low-carbon-economy-in-the-information-age/> Retrieved 08.08.11.

UN (2009). UN department of economic and social affairs. Division for sustainable development. Sustainable development topics. <http://www.un.org/esa/dsd/susdevtopics/sdt_index.shtml> Retrieved 08.08.11.

UNFCCC (2011). Kyoto protocol. <http://unfccc.int/kyoto_protocol/items/2830.php> Retrieved 12.08.11.

Vaquero, L. M., Rodero-Merino, L., Cáceres, J., & Lindner, M. (2009). A break in the clouds: Towards a cloud definition. *ACM Computer Communication Reviews, 39*(1), 50–55.

Wiedmann, T., & Minx, J. (2008). A definition of 'carbon footprint'. In C. C. Pertsova (Ed.), *Ecological economics research trends* (pp. 1–11). Hauppauge NY: Nova Science Publishers, Inc.

WSRCC at ECOOP 2011. *3rd International workshop on software research and climate change*, <http://sites.google.com/site/wsrcc2011> Retrieved 08.08.11.

Yang, Z., Kamata, S.-I., Ahrary, A. (2009). NIR: Content based image retrieval on cloud computing. In *2009 IEEE international conference on intelligent computing and intelligent systems, ICIS 2009*; Shanghai; 20 November 2009 through 22 November 2009, Proceedings-2009 IEEE international conference on intelligent computing and intelligent systems, ICIS 2009 (Vol. 3, pp. 556–559), Article number 5358101.