



GLOBAL JOURNAL OF COMPUTER SCIENCE AND TECHNOLOGY
SOFTWARE & DATA ENGINEERING

Volume 12 Issue 12 Version 1.0 Year 2012

Type: Double Blind Peer Reviewed International Research Journal

Publisher: Global Journals Inc. (USA)

Online ISSN: 0975-4172 & Print ISSN: 0975-4350

Intelligent Information Retrieval

By Ayushi Gupta, Divya Verma & Kajal Gambhir

Dronacharya College of Engineering

Abstract - The World Wide Web has become an invaluable information resource but the explosion of information available via the web has made web search a time consuming and complex process. Index-based search engines, such as AltaVista, Google or Info seek help, but they are not enough. This paper describes the rationale, architecture, and implementation of a next generation information gathering system – a system that integrates several areas of Artificial Intelligence (AI) research under a single umbrella. Our solution to the information explosion is an information gathering agent, IIR, that plans to gather information to support a decision process, reasons about the resource trade-offs of different possible gathering approaches, extracts information from both unstructured and structured documents, and uses the extracted information to refine its search and processing activities.

GJCST-C Classification: H.3.3



Strictly as per the compliance and regulations of:



Intelligent Information Retrieval

Ayushi Gupta^α, Divya Verma^σ & Kajal Gambhir^ρ

Abstract - The World Wide Web has become an invaluable information resource but the explosion of information available via the web has made web search a time consuming and complex process. Index-based search engines, such as AltaVista, Google or Info seek help, but they are not enough. This paper describes the rationale, architecture, and implementation of a next generation information gathering system – a system that integrates several areas of Artificial Intelligence (AI) research under a single umbrella.

Our solution to the information explosion is an information gathering agent, IIR, that plans to gather information to support a decision process, reasons about the resource trade-offs of different possible gathering approaches, extracts information from both unstructured and structured documents, and uses the extracted information to refine its search and processing activities.

I. INTRODUCTION

The World Wide Web has given the researchers, businessmen, corporate, students, hobbyists and technical groups a medium by which they can share the information they have, with others. The ease of HTML and platform independence of the web documents has led to a tremendous growth of the web, that has outstripped the technologies that are used to effectively search in these pages, as well as proper navigation and interpretation.

With the aim of inception of AI (Artificial Intelligence) in the searching techniques, the first step we have decided is to find out those limitations in the current searching methodologies, which make the result unsatisfactory and not up to the expectations. Some of the key features of today's search engines are:

- **Meta Searching:** The scope of each search engine is limited and no search engine has the database that covers all the web pages. This problem was noted long ago and was solved with the help of Meta search sites that make use of multiple search engines to search for the "Query String. The common names of such search engines are 37.com (which searches 37 search sites simultaneously), metacrawler.com and many others. Another advantage of these Meta search sites is that they incorporate advanced features which are absent in some of the member search sites (Member search sites are those sites which return the search result to Meta Search engines). But the basic methods used in these Meta search sites are more or less same as those used in any other search engines.

- **URL Clustering:** URL clustering was a basic problem from which most of the earlier search sites were affected. Suppose we search for 'GRE' and we intend to get the link to all those sites that have information on GRE exam. But a search engine without URL clustering will give results.
- **Shopping Agent:** It is an intelligent enhancement over the other searching techniques, which *tries* to give the most appropriate site, not just any site that has high frequency of the Key Phrase. For example, if a person with the intention of shopping searches the web for 'Printer', then the normal search engines will return the pages which have high frequency of the word - 'Printers'. There can be a case when one of the results contain the information irrelevant from the point of shopping, but still that page has very high frequency of Query String. Say, a person on his personal home page writes : 'My **printer** HPDeskJet 640C is a 1997 model', then the search engines will return this page as well (Note that the user has not used any Boolean Operator in the search string). While a shopping agent gives the details in a commercial format like price range, model, other options, second hand options and just many results. The implementation of Shopping Agent was first serious step forward in the direction of making *Intelligent Information Retrievers*. We will be using its powers in the design of our new Information Gathering Agent.
- **Personal Information Agent (PIA):** This is the most important step forward in the direction of incorporating AI in searching. The PIAs try to retrieve your personal interests and give the result accordingly. The information is gathered either from the previous search results and mostly by a questionnaire. But the current day PIAs are very slow and less adapting. In this paper, we will try to confer PIA with the power to give satisfying and fast search results.

II. INTELLIGENT INFORMATION RETRIEVAL

The solution to the problem of *Intelligent Information Retrieval* is to integrate different Artificial Intelligence (AI) technologies, namely scheduling, planning, text processing, information extraction and interpretation into a single information gathering agent, which we christen as **Intelligent Information Retriever (IIR)**. IIR locates, retrieves and processes information to support a human decision process. During thinking, we human adopt a top-down and a down-top structured

E-mail ^α : sweetuaayushi@gmail.com
E-mail ^σ : divyaverma417@gmail.com
E-mail ^ρ : kjlgambhir7@gmail.com

analysis. Before discussing how this can be implemented through AI, first let's have a glimpse at how human is able to do this. For this, we create a scenario in which a person wants to buy a book that is not available at the local book stores. The person now has two options: Order the book from the publisher and second option is to go to a nearby town and have the book from there, provided that the person has the information that the book is available at the book stores of that city. To complicate the situation, further assume that the book is by a foreign publisher and that publisher has no branch in the country of the person, so ordering a book from the publisher will result in a time consuming process. Let us further assume that the overhead expenses involved in visiting the neighboring town is more than the actual cost of the book. Now the person will subconsciously list all the parameters in his mind that may affect the decision of buying the book. The typical, probably minimum list of questions that will come in his mind are:

1. Whether the book really worth buying?
2. Whether the book is required urgently?
3. Is there any alternative to that book?
4. Do I have enough money to buy that book?
5. Do I have enough money to bear the overhead expenses involved in visiting neighboring town/city?
6. How will I get to the neighboring city / How will I order the book from the publisher?

So, in any such decision making, humans make use of following :

- Interpretation [derived from pt. 1 and 2 above]
- Comparison [pt. 3]
- Monetary Factors [pt. 4 & 5]
- Planning and scheduling [pt.6].

Our aim is to incorporate above decision making constructs in searching making proper use of AI (Artificial Intelligence). We will be implementing all this through a new information-gathering agent, that we have already christened as Intelligent Information Retriever (IIR).

"The IIR is a data-driven as well as expectation-driven and adaptable information gathering agent that does information interpretation through decision making constructs properly adjusted and incorporated with the existing powerful points of today's search engines, most prominent of which being Personal Information Agent and Shopping Agent." After having formally designed the definition of IIR, we are in a position to be equipped with the tools and techniques that will be used in the design of IIR.

III. CRITICISM

The integration of different components in IIR - The Task Assessor, Decision Maker, CORE, Object database, Information Extractor is itself a major

accomplishment in its own kind. Despite the integration issues, the combination of the different AI components in IIR and the view of information gathering as an interpretation task have given IIR some very strong abilities.

In terms of limitations, the following points should be noted:

- Initially, due to smaller Object Database, the results will be lesser efficient (but still more efficient than current technology). This problem can be overcome by having a large database before the start of the service.
- The form fields to be filled by the user *may* increase, if precise results are desired.
- The cost of implementation will be very high.

Despite these limitations, this Intelligent Information Retriever is a major enhancement over the current search engines and is a serious step forward in the direction of incorporating Artificial Intelligence in searching for more efficient results.

REFERENCES RÉFÉRENCES REFERENCIAS

1. Intelligent information retrieval. en.wikipedia.org/wiki/intelligent_information_retrieval
2. Intelligent techniques for IIR www.sigir.org/forum/2006D/2006d_sigirforum_siddiqui.pdf
3. Intelligent information retrieval laboratory iir.csie.ncku.edu.tw