ISBE Imaging Science and Biomedical Engineering **IMAGING SCIENCE AND BIOMEDICAL ENGINEERING** MANCHESTER, UK M13 9PT

IURON: SUCCINCT FORMAL PROPOSITION

ROY S. SCHESTOWITZ
IMAGING SCIENCE AND BIOMEDICAL ENGINEERING
UNIVERSITY OF MANCHESTER
UNITED KINGDOM

Version 1.0: An initial project overview

Abstract

Search engine technologies have been discussed to death and further developed endlessly in the past decade. However, such engines have no so-called "thirst for knowledge", but rather a thirst for text. We continue to live in an age where best results for a query are produced given an input comprising *keywords*. The outcome, rather than answers or self-tailored content, is merely a linear collection of pages, whose static content *resembles* the keywords. There is no way to guarantee, nonetheless, that such pages will provide the desired information or provide information that is credible.

Iuron is set to become a collection of tools for *knowledge* engines, which are intended to crawl the World Wide Web. The aim is to create a semantic entity that captures *facts* from a large number of pages, thereby providing an *intelligent* front-end for user search. Results are generated 'on the fly' based on acquired knowledge and are solely intended to serve individual users.

Overview

LET us think of the Internet as a collection of complex, inter-related information. More cohesively, it takes an immense number of hypotheses and thus can contain valid, consistent knowledge. Although we can process (scan) all the information, higher-level knowledge, which is derived from *collection* of pages, is still missing. There is enough knowledge across the World Wide Web to answer more or less any question, assuming it is not subjective. All that is done at present is word indexing with the notion of work proximity.

Let us face the fact that, among the more popular uses of search engines, are pursuits for commercial companies, which provide products or services. Results that get returned by the engines sometimes correspond to the most valid and relevant *authority* for a given niche. This may be fine for insight into magnitude and breadth of companies (or their Web sites), but this equally often misleads the user.

Search engines at present fail to extend beyond a potentially morbid state of "dominance prevails". Rather than an engine that provides users with the *most reasonable* answer and/or reference to a site, it provides a Web link to what is most cited, typically due to fraudulent practices or subjective search engine optimisations.

All the all, search engines at present encourage link-related spam and content-related spam. In worse scenarios, their backlinks-based algorithms lead to rise in sponsored listings, whereas our natural incentive is to prefer what would *"work best for us"*, not what got recommended by automated tools. These tools, which work at a shallow level without *understanding*, opt to prioritise large corporations with money to be spent on good listings and inbound links.

Iuron is a project that addresses the issues above. First and foremost, it converts the vast amount of information in the World Wide Web into facts. Moreover, it serves as an impartial source for answers and is not highly susceptible to deceit as it can discern true from false.

Methods

There are a variety of plausible ideas, which have been expressed at some depth in the manifestation document alongside their pitfalls. To name one of them briefly, pages should be obtained from the World Wide Web and then reduced to a set of facts. Facts will be assigned varying weighs depending on credibility factors. Frequently-repeated facts will be encouraged while falsified facts discouraged or altogether rejected. First-order logic serves as the holy grail by which a sequence of words (elements) becomes a set of arguments with associated semantics.

Practicability

The fundamental approach to tacking the problem is not overly complicated. The goal is certainly feasible, while the resources to make it practical are the primary barrier.

Since Iuron is an Open Source project, rapid assemblage and construction of the libraries would be rapid, making use of existing projects that fall under the General Public Licence (GPL). In return, Iuron will provide a potentially distributed environment, wherein any idle computer across the world can assist crawling and report back to a main knowledge repository. Think of it as a public-driven reciprocal effort to process and then centralise human knowledge.