

Accepted Manuscript

A semi-automated framework for semantically annotating web content

Mohamed Abdou, Sayed Abdel Gaber, Marwa Farhan

PII: S0167-739X(17)30734-3
DOI: <https://doi.org/10.1016/j.future.2017.11.008>
Reference: FUTURE 3800

To appear in: *Future Generation Computer Systems*

Received date: 24 April 2017
Revised date: 2 November 2017
Accepted date: 5 November 2017

Please cite this article as: M. Abdou, S.A. Gaber, M. Farhan, A semi-automated framework for semantically annotating web content, *Future Generation Computer Systems* (2017), <https://doi.org/10.1016/j.future.2017.11.008>

This is a PDF file of an unedited manuscript that has been accepted for publication. As a service to our customers we are providing this early version of the manuscript. The manuscript will undergo copyediting, typesetting, and review of the resulting proof before it is published in its final form. Please note that during the production process errors may be discovered which could affect the content, and all legal disclaimers that apply to the journal pertain.



A Semi-Automated Framework for Semantically Annotating Web Content

Mohamed Abdou¹, Sayed AbdelGaber², and Marwa Farhan³

Information Systems Department,

Faculty of Computers and Information, Helwan University, Egypt

¹mabdou@fci.helwan.edu.eg, ²sgaber@fci.helwan.edu.eg, ³marwa.salah@fci.helwan.edu.eg

Abstract

Today's web is growing very fast and having a strong online presence is becoming critical for businesses. In contrast, websites in search results using traditional Search Engine Optimization (SEO) techniques became less effective in achieving the desired visibility. To overcome the limitations of traditional SEO, some recent trends are adopting the technologies of Semantic Web to annotate web content with Semantic Markup that can be understood by search engines. However, the balance between the accuracy of annotations and the automation level still under investigation. This research proposes a semi-automated framework that provides a high level of accurate annotations with minimum user interaction, based on Schema.org; a well-accepted ontology for ordinary things in life. The proposed framework aims to analyze the contents of web documents and extract the unique keywords and key phrases that best describe that content. Then, annotate those keywords and key phrases with the appropriate Schema.org vocabularies. This will reflect on the understanding level of the search engines to the web contents and accordingly a better visibility in the search results.

Keywords: Semantic Markup; semantic annotation; Schema.org; JSON-LD.

1. Introduction

The search engines, in general, do not understand the meaning of information within web pages, as their contents created for humans, not for machines. The search engines need to understand what the exact meaning of the search query is. Moreover, they need to determine which web pages contain the appropriate information [1]. Traditional websites contents are mainly human readable only, which is inefficient for automatic processing to search for requested information. In contrast to that traditional Web, the Semantic Web is machine readable; it refers to the web of data, which means it connects things rather than documents that have no meaning to computers [2,3].

In a featured article in Scientific American, TimBerners-Lee et al. (2001) [4] introduced the Semantic Web as "an extension of the current Web, in which information is given well-defined meaning, better-enabling computers and people to work in cooperation." They believed that Semantic Web would help in structuring the meaningful content of web pages, and facilitating the work of software agents as they move from page to another to carry out complicated tasks. Accordingly, machines do not just display the data, but they would be able to process and understand this data.

Recently, the major search engines made updates to their algorithms to get the meaning of users' search queries. So, it is necessary to update SEO techniques to cope with those changes [5,6]. The Semantic Web evolution offered a solution for such issue through adding semantics to web pages, which makes them readable for humans and machines [1,7]. The Semantic Web technologies make it possible to add Semantic Markup to web pages through annotating their contents. Consequently, there were several attempts to include

Download English Version:

<https://daneshyari.com/en/article/6873258>

Download Persian Version:

<https://daneshyari.com/article/6873258>

[Daneshyari.com](https://daneshyari.com)