Journal of Theoretical and Applied Information Technology
31st January 2019. Vol.97. No 2

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

SEMANTIC INFORMATION EXTRACTION APPROACH FOR E-COMMERCE SEARCH ENGINE BASED ON GOODRELATIONS ONTOLOGY

[1]ABDELHADI BAHAFID, [2] KAMAL EL GUEMMAT, [3] EL HABIB BEN LAHMAR, [4] MOHAMED TALEA

[1] PhD candidate of Mathematics, Informatics and Information Processing in the faculty of science Ben M'sik, Hassan II University, Mohammedia-Casablanca, Morocco,
[2]PhD in computer sciences, Research fields: Semantic Indexing, Semantic Web, Information Retrieval Systems, Automatic Processing of Natural Language, Information and communication technology for teaching and learning, Educational modeling, E-Learning,
[3,4] Professor of Higher Education at the Faculty of Sciences Ben Sik, Casablanca, Morocco

E-mail: [1]bahafidabdelhadi@gmail.com, [2]h.benlahmer@gmail.com, [3]k.elguemmat@gmail.com, [4] taleamohamed@gmail.com

ABSTRACT

Internet for e-commerce is the main source of information, this information is not directly exploitable by computers, hence many methods and approaches to extract this information, in order to use them. Search engines [1] use these methods or approaches to extract and index the information contained in the web pages.

Users use search engines to find useful information about the products they need, which shows the importance of search engines and having to equip them with good extraction methods to respond more accurately and in a relevant way to the need of users.

Most of these search engines are based on keywords [2] to extract and index data from web pages, which explains the quality of the search results [3] of these engines which often return results that does not match the search performed, the result is not always relevant, hence the approach proposed in this article, it is a new approach that consists of linking the CSS incorporated on the e-commerce web page and GOODRELATIONS ontology used to index these web pages by means of a database, and from these CSS classes, generate a Wrapper to extract all the information about the products, which allow us to know the attribute of each product, it corresponds to which attribute of the ontology, i.e. its semantics, this will improve the relevance of the results of the research and respond more precisely to the need of the user.

Keywords: *Web Semantic, Information Extraction, Information Retrieval System, Semantic Indexing, E-Commerce, Ontology, GOODRELATIONS*

1. INTRODUCTION

The web has become rich in information circulating throughout the world via the internet network. This caused to the expansion of large amounts of data, and these data are often unstructured and difficult to use by web applications. However this amount of data cannot be easily accessed and manipulated, the main reason is the textual nature of web content that makes Web data sources are intended to be covered by users (humans) and not by the Machines, hence the complexity of the task of extracting data from the web, this task is a crucial task in an Information Retrieval System, it allows to extract the data, i.e. the concepts and that data are indexed afterwards in order to be able to answer the requests of the users, hence the need to have a good method of extraction.

The problem "How to access to the information available in the form of HTML pages?" Is still relevant, at least in the short term, and represents an enormous challenge for the years to come. The design and development of data retrieval systems remains a highly needed, which explains that search

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

engines cannot satisfy the needs of users because of their low recall and precision, because most search engines are based on keyword.

There is a semantic gap between keywords and concepts, for example, the same keywords may have different meanings in different contexts, and therefore the returned results match only the user's query with words rather than concepts.

Especially on e-commerce, the search for product information, current search engines based on keywords have a low recall rate due to the fact that the same product may have different classifications, different names and different literal descriptions. So, when a customer wants to buy a certain product on the web, he has to browse many websites as he can to get appropriate information about the desired product characteristics, attributes, performance, prices and so on.

Obviously, it is not conducive for customers who want to buy on the internet, to do such boring and lengthy work, hence the need to look for other approaches to extract the data as well as their semantics, so that the answer to the user requests are more relevant and that will facilitate to the search engines acquiring those information used for users response.

Several approaches have been proposed: “Gene / Clone approach, the structural prefix and structural suffix” [4] it needs a great manual effort and it has no semantic contribution, “Knowledge-based Wrapper Generation by Using XML” [6] it only works for labeled documents, not functioning on table-type documents, ”A Shopping Agent That Automatically Constructs Wrappers for Semi-Structured Online Vendors" [7] this approach is based on keyword, ….

In this sense of improving the extraction methods that our new extraction method is registered which will allow us to extract the embedded data in the e-commerce pages as well as their semantics by making the link between the CSS classes used in these pages and the attributes of GOODRELATIONS ontology.

Certainly, in the middle term, migration to XML format should allow easier access to this information because this language not only makes documents readable but also exploitable and usable by machines.

Our goal is to propose semantic information retrieval system to help the user to acquire the desired information on desired products and address the problems mentioned above to improve the relevance of IRS (Information retrieval system) by adopting semantic web technologies. From the set of responses obtained from the IRS, we can measure the relevance of this one. The criteria for measuring relevance are recall and precision.

This article will be divided as follows: we define in the next section some web extraction methods, the subsequent section presents our contribution as well as our information retrieval system, and the last section presents conclusion and perspectives.

2. STATE OF THE ART

The main source of information is the World Wide Web, One of its main great expansion is the absence of any strict rules for the information presentation and the relative simplicity of the used technology. HTML (Hypertext Markup Language) is the main programming language used to create documents, it gives the authors enough freedom for presenting any kind of data with minimal effort and the new technologies such as Cascading Style Sheets (CSS) allow to achieve the desired quality of presentation. The different forms of data presentation on the web, also have a lot of drawbacks, it increases the complexity of how to access, extract and use effectively all the information contained in the web. Since the majority of search engines use HTML, it explains the need to have good extraction methods to have an IRS that responds appropriately to the needs of users.

2.1 Information Extraction from HTML Documents

The information extraction has been largely investigated in the plain text context – it has been used for processing electronic mail messages, network news archives etc. [8]. The HTML language brings a new look to this problem. In contrast to the plain text messages, HTML allows to define the visual presentation of the content. This possibility is often used for making the documents clearer and easily understandable. The data in HTML documents is presented in a more or less regular and structured fashion. For this reason, the HTML documents are often regarded a semi structured information resource [9]. The reader is not forced to read the whole document, he can easily find the part of the document containing the desired information. For this purpose, the documents contain a system of navigational cues that have mostly visual character. During the years of the World Wide Web development, these

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

techniques of data presentation have been brought almost to perfection so that reading the documents using the web browser becomes relatively efficient.

From the point of view of the automated document processing, the situation is different. HTML doesn't contain sufficient means for the machine-readable semantic description of the document content. The techniques for natural language processing [10] that have been used for the information extraction from plain text are not applicable, because HTML documents usually don't contain many whole sentences or blocks of continuous text. On the other hand the HTML tags inserted to the text of the document provide additional information that can be used for identifying the data.

For the data in the HTML documents, a database terminology is usually used in the information extraction context. We assume that the documents contain one or more data records where each record consists of some number of data fields. Usually, we admit that some records are incomplete; i.e. that the values of some fields are not present in the document.

2.2 Information retrieval

Finding relevant information [11] within a large collection of documents is one of the first challenges information systems had to face. For that a set of computer programs and systems were developed, but it was not until the beginning of the Web that most important developments came about.

Each user tries to locate documents that can yield information that he requires i.e. each user tries to satisfy his information needs. The process of identifying (cf. figure1), searching and acquiring the potential documents that may meet these information needs is called user retrieval process [12]. All of the retrieved documents aim at satisfying user information needs expressed in natural language.

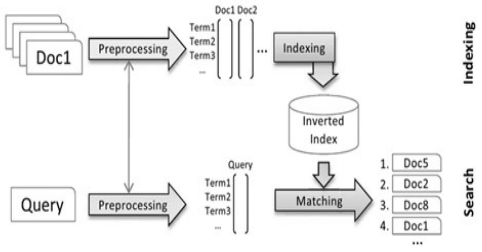

Figure 1: Information retrieval process

A part of computer science that studies the retrieval of information from a collection of written documents is called Information retrieval (IR) [13].

Information retrieval (IR) (cf. figure2) consists in obtaining relevant information from a collection of information resources. Searches can be based on full-text or other content-based indexing. Information retrieval is the science of searching for information in a document, searching for documents themselves, and also searching for metadata that describe data, and for databases of texts, images or sounds.

Automated information retrieval systems are used to reduce what has been called information overload. Many universities and public libraries use IR systems to provide access to books, journals and other documents. Web search engines are the most visible IR applications.

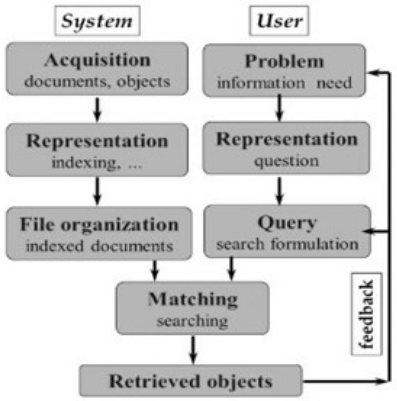

Figure 2 : Traditionnel information retrieval system

2.3 Extraction approaches:

In the literature a lot of work were done to develop new extraction approaches, to improve IRS classical approaches: "Gene / Clone approach, the structural prefix and structural suffix", "Knowledge-based Wrapper Generation by Using XML" , "A Shopping Agent That Automatically Constructs Wrappers for Semi-Structured Online Vendors", the majority of these approaches are based on keywords, i.e. all extracted information has a textual nature and we do not have their semantic, there is always a semantic gap between keyword and concepts. and it's reflected in the search engine result, which is not always very accurate in answering user queries. …. On the other hand semantic approaches, we conducted from our

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

research carried out their rarity example: "Semantic information extraction for software requirements using semantic role labeling" [14], and even less the extraction approaches dedicated to e-commerce. This rarity of work around e-commerce, which occupies an important place in the life of the users of the web, motivated us to dig the subject and make this study which gave birth to this approach.

we will propose in this article an approach that allows the extraction of information about products, it consists of linking the CSS incorporated on the e-commerce web page and GOODRELATIONS ontology used to index these web pages by means of a database. The next section explains with more details our approach.

3. CSS extraction approach

The aim of our approach is to identify a particular information that is explicitly contained in the text of the e-commerce web pages and to store it in a structured form e.g. to a database table or a XML document, it allow us to extract just the useful information used to index the document and to know the attribute of each product, it corresponds to which attribute of the ontology, i.e. its semantic and that will help us to respond to request user with more precisely.

3.1 Block and element

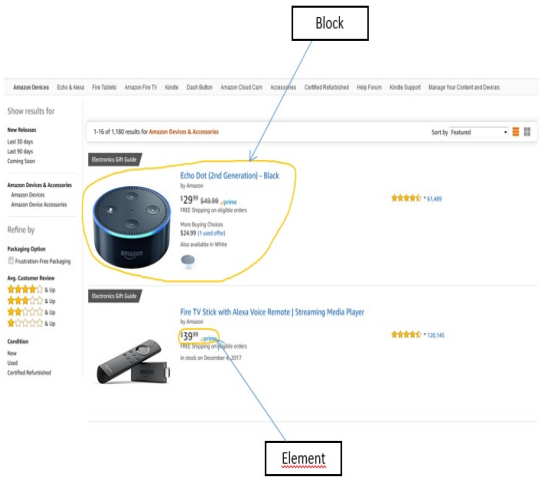

Figure 3: Block and Element

A block (cf. figure 3) is an independent entity, a brick of the web page. A block forms its own context.

An element is a part of a block. The context of an element is that of the block.

3.2 Extraction approach

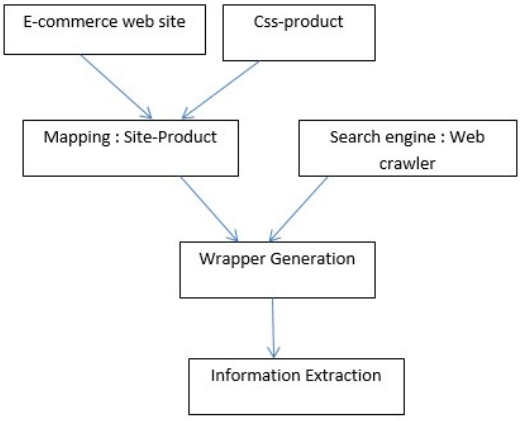

Figure 4: Extraction approach

By consulting some e-commerce sites (walmart, Amazon, CDISCOUNT, laredoute.fr ...), we notice that they generally display the products in the same way: a set of block, each block describes a product. We will use this statement to propose a new approach that allows extracting all relevant information embedded in an e-commerce web page about products.

The e-commerce sites are composed of blocks which are themselves composed of elements, all the blocks that represent the products have the same rendering and use the same CSS classes to describe different products and the elements contained in these blocks use the same CSS classes too.

This extraction approach consists in making the link between the embedded CSS on the e-commerce web page and the GOODRELATIONS ontology used to index these web pages through a database.

The Database will contain for each site and for each attribute of a product (element) in this site the CSS class that describes it as well as the corresponding attribute in the GOODRELATIONS ontology.

The database is used by the search engine, so that for each page embedded in a site, it will retrieve the list of classes that represent the set of attribute's a product in this site and generates a Wrapper to extract all products (information about each

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

product) and this information that will use to index all the web pages of the site in question.

3.2.2 Fill in the Database

Filling the database (cf. figure 5) is a crucial phase allows to make link between e-commerce website and GOODRELATIONS ontology, it allows for each e-commerce site to insert the CSS classes that describe the products on the site and their correspondents in the ontology, we have a kind of ontology mapping-site.

By accomplishing this phase, we can know for each attribute in e-commerce page corresponds to which attribute in the ontology GOODRELATION, so that the search engine knows in advance before browsing the page each field in this page and its equivalent in the ontology i.e. the field and its semantics and once it extracts these information it can save them directly to build its index.

Example:

There are several ways of doing and designing for this part, we present here one of many others, below (cf Table 1, Table 2, Table 3, Table 4) the tables and links between tables that can be used to implement our approach.:

Code_Site	Name	Link
AMZ1	Amazone	https://www.amazon.fr/

Table 1: "Site"

Id_Attribute SITE	Attribute name
1	Name
2	Category

Table 2: "Site Attributes"

Id_Attribute ONTOLOGY	Attribute name
1	Name
2	Category

Table 3: "GOODRELATIONS_PRODUCT_INFO"

Id_Link	Code_Site	Id_Attribute_Site	Id_Attribute_ONTOLOGY
1	AMZ1	1	1
2	AMZ1	2	2

Table 4: "LINK_SITE_ONTOLOGY"

Initially, this phase will be done manually, in order to improve in the next publication, the figure below (cf shows how database will be filled:

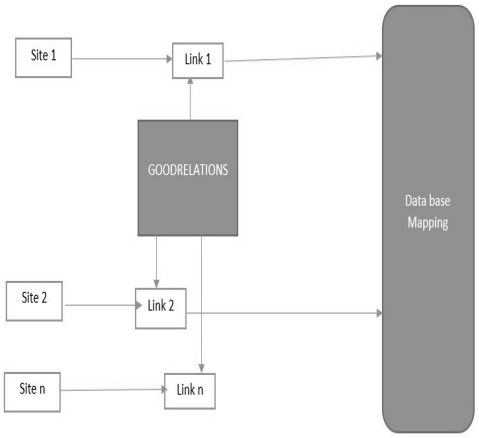

Figure 5: Filling the database

3.2.3 Data extraction:

Our aim in the HTML e-commerce pages processing is to identify a particular information that is contained in the text of pages and to store it in a structured form (database, XML file, OWL file,). This process is called information extraction from HTML document.

To apply our approach and extract information from an e-commerce site, the search engine starts by consulting its Mapping database, to extract the different CSS classes that represent each product in the e-commerce page, once this information is obtained it generates a wrapper from these classes to extract all the information about all the products that the e-commerce page contains, using algorithm above:

DECLARE
//information about site to parse
site: Site;

/* represents the CSS class of the element (often a div) that contains the product information*/
node_css: Product_CSS_node;

Page: DOC_HTML;

productsBrutes: String;
attributes: Product Attributes;
//values extracted from the page

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

```
attributesValues : values of the product attributes;
/* object that contains the extracted attribute
values*/
index: ObjectToIndex;
begin
 // read the content of the page
  load (page);

/* read the node of product (root of each product),
value saved in the database*/
node_css <- readNode (site);

/* Retrieve the different attributes (that we already
registered in the database) of the product contained
in this page (the concerned site) */
attributes <- readAttributes(site);

/*
Retrieve all products contained in the page using
the CSS class "Product_CSS_node" (class = value
read from the database)
*/
productsBrutes <-readProducts(page,node_css);

For each product in productsBrutes

attributesValues <-
extractProductAttributeValue(produit,attributes);

/*
feed the object from the attributes of the current
product
*/
 index <- alimenterObjet(attributesValues);

writeInIndex (index); // Save the index object

END For

end
```

The result of this phase is a set of information about products contained in the web page, this set of information that will be indexed to respond to user requests.
The figure below (cf. figure 6) shows the process of extracting information.

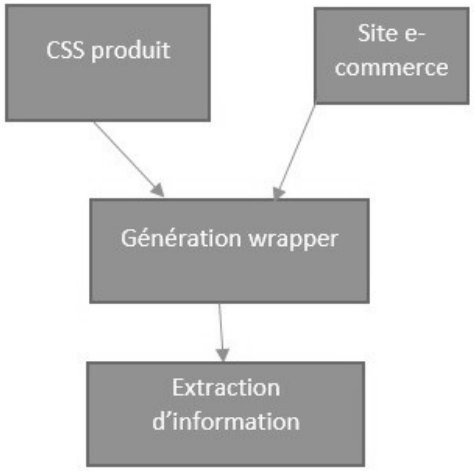

Figure 6: Extraction process

3.2.4 Semantic indexing

The semantic indexing [16] will allow us to represent the various product information.
The semantic indexing includes two steps [17]:

Disambiguation: get the exact meaning of words extracted from the document to be indexed.
Representation: to represent the document in order to retrieve information.

3.2.5 Disambiguation

The extraction phase allows to extract information in an unambiguous way, because for each extracted attribute, it is known what information is related to the product, as well as which attribute in the ontology, so this phase is included in the extraction phase.

3.2.6 Representation

To represent the document, we will use as a model the ontology GOODRELATION.

Ontology was originally a concept of philosophy to study being, i.e. the study of general properties of what exists. And later the artificial intelligence researchers reused concept for modeling domain knowledge.

An ontology includes concepts that represent all knowledge of a domain in an explicit and formal specification ([18]). It shows the relationships and rule's association between these concepts to allow

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

the one hand to computers the generation of new knowledge through inference, and secondly to users and computer to give common meaning to the terms used in a business area in order to remove any ambiguity during treatments.

3.2.7 GOODRELATIONS Ontology

GOODRELATIONS [19] is a web ontology for e-commerce, is the most powerful vocabulary for publishing all of the details of products and services in a way friendly to search engines, mobile applications, and browser extensions. By adding a bit of extra code to the Web content, it make sure that potential customers realize all the great features and services and the benefits of doing business with vendors, because their computers can extract and present this information with ease.

1. Conceptual Overview

The goal of GOODRELATIONS is to define a data structure for e-commerce that is

- **industry-neutral**, i.e. suited for consumer electronics, cars, tickets, real estate, labor, services, or any other type of goods.
- **valid across the different stages of the value chain**, i.e. from raw materials through retail to after-sales services.
- **syntax-neutral**, i.e. it should work in microdata, RDFa, RDF/XML, Turtle, JSON, OData, GData, or any other popular syntax.

This is achieved by using just four entities for representing e-commerce scenarios:

1. An agent (e.g. a person or an organization),
2. An object (e.g. a camcorder, a house, a car,..) or service (e.g. a haircut),
3. A promise (offer) to transfer some rights (ownership, temporary usage, a certain license, ...) on the object or to provide the service for a certain compensation (e.g. an amount of money), made by the agent and related to the object or service
4. A location from which this offer is available (e.g. a store, a bus stop, a gas station...).

This Agent-Promise-Object Principle can be found across most industries and is the foundation of the generic power of GoodRelations. It allows you to use the same vocabulary for offering a camcorder as for a manicure service or for the disposal of used cars.

The respective classes in GOODRELATIONS are:

- **gr:BusinessEntity** for the agent, i.e. the company or individual,
- **gr:Offering** for an offer to sell, repair, lease something, or to express interest in such an offer,
- **gr:ProductOrService** for the object or service, and
- **gr:Location** for a store or location from which the offer is available (in previous versions: **gr:LocationOfSalesOrServiceProvisioning**).

2. UML Class Diagram

The following UML class diagram (*Figure 7)* illustrates the interplay between the GOODRELATIONS language elements. Please note that UML class diagrams are only approximations of OWL ontologies and microdata vocabularies.

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

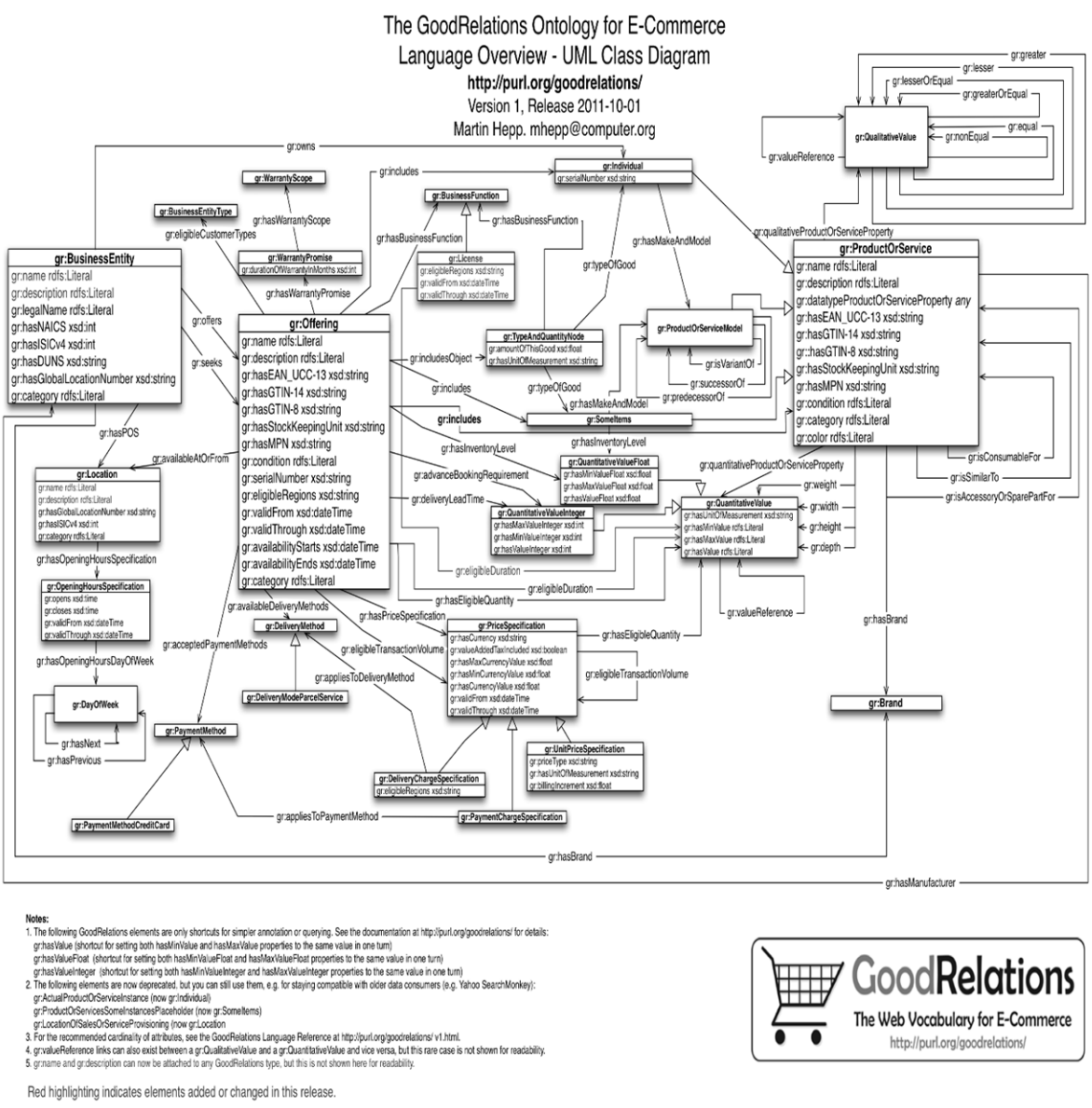

Figure 7: UML class diagram

4. ADOPTION OF THE EXTRACTION APPROACH FOR AN E-COMMERCE SEARCH ENGINE

Information retrieval is the science that meets the needs of users to select the most relevant results. In order to accomplish this task we need an Information Retrieval System (IRS). An Information Retrieval System is an interface between users who express their needs through queries and a large collection of documents. The information retrieval system processes these documents to display to the user the documents or products that best match his request. To do this, it integrates a set of functions which are the main ones: storage, organization and selection of information that meet the needs of users and finally the return of information deemed relevant. The purpose of the information search (cf. figure 8) is to satisfy a user's need for information, this need is expressed by a request, on a set of documents called a collection or a corpus. An information

Journal of Theoretical and Applied Information Technology
31st January 2019. Vol.97. No 2

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

retrieval system is defined by its models of document representation, user needs, and its matching function.

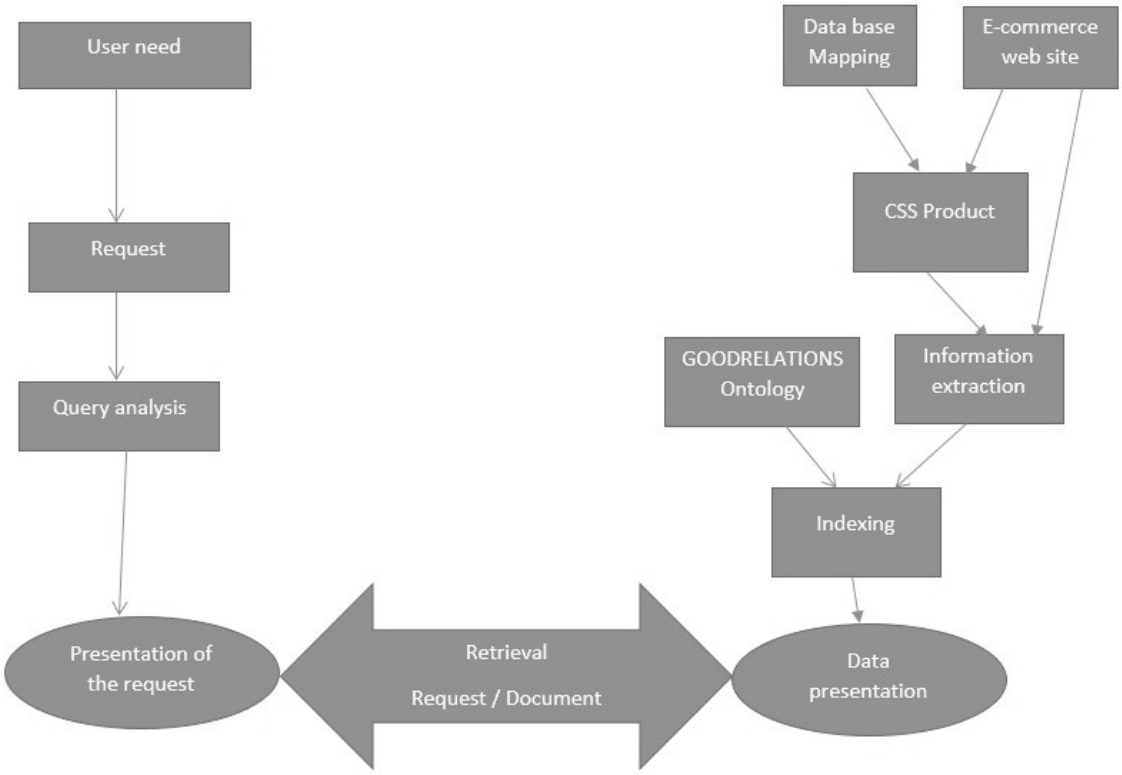

Figure 8: Information Retrieval Architecture

5. EXPECTED IMPLEMENTATION :

The implementation of this approach is expected to use Oracle and StormCrawler.

Oracle:
We use oracle as database, to store for each e-commerce site the corresponding CSS classes that describe the products on the site and their correspondents in the ontology.

StormCrawler:
StormCrawler is an open-source collection of resources for building low-latency, scalable web crawlers on Apache Storm. It is provided under Apache License and is written mostly in Java (programming language).
The aim of StormCrawler is to help build web crawlers that are:

- scalable
- resilient
- low latency
- easy to extend
- polite yet efficient

StormCrawler is a library and collection of resources that developers can leverage to build their own crawlers.
Once the crawler is developed, it can be launched to retrieve data from different e-commerce sites, build the index so that we can respond to users' requests.

6. EVALUATION OF OUR NEW E-COMMERCE SEARCH ENGINE:

The evaluation of our approach consists of measuring two quantities and one combined measure. The 'Recall' and 'Precision' rates are the two quantities, and the F-measure is the combined measure. The Recall, Precision, and F-measure are calculated as the followings:

$$\text{precision} = \frac{|\{\text{relevant documents}\} \cap \{\text{retrieved documents}\}|}{|\{\text{retrieved documents}\}|}$$

$$\text{recall} = \frac{|\{\text{relevant documents}\} \cap \{\text{retrieved documents}\}|}{|\{\text{relevant documents}\}|}$$

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

To evaluate this approach, we used a page (https://www.cdiscount.com/ and we search 'montre guess collection femme', calculation performed 2 October 2018) with a hundred words and products on which we applied our approach. The table (cf. figure 9) below show the result:

Products extracted by search engine and which are accurate according to the expert	Noise (False products extracted by search engine acco rding to the expert)	Silence (products not extracted by search engine acco rding to the expert)	Precision	Recall	F-mesure
51	0	8	1	0.85	0.92

Figure 9: Precision and recall

The precision of our approach shows that the result found is relevant, the recall is good and shows that it exist some products that we have not found and it is due to the fact that the page has some products that do not respect the same format as the others and it is not very current which shows that the result obtained is relevant.

7. CONCLUSION AND PERSPECTIVES

Proceeding from general considerations about the nature of IE problem we have proposed a new approach of extracting information from e-commerce web page based on GOODRELATIONS ontology and data base, a new information retrieval system based on approach we have proposed, to help search engines acquiring relevant information about products and then satisfy the needs of users to acquire the desired information on desired products and the value of the recall and precision shows that our approach returns relevant results.
Several perspectives are possible for this approach, first automating the management of site-ontology mapping to avoid doing it manually, implementing the information retrieval system and Apply this approach to other domains such as the medical field, artistic ...

REFRENCES:

[1] Croft, W. B., Metzler, D., & Strohman, T. (2010). Search engines: Information retrieval in practice (Vol. 283). Reading: Addison-Wesley.

[2] Diamond, Ted, et al. "Keyword-based search engine results using enhanced query strategies." U.S. Patent No. 8,645,372. 4 Feb. 2014.

[3] Chu, Hung-Chi, et al. "The Improvement of Web Page Ranking on SERPs." 2018 IEEE International Conference on Consumer Electronics-Taiwan (ICCE-TW). IEEE, 2018.

[4] Benlahmer EL Habib,Doukkali Seduigi Aziz,El ouerkhaoui Asmaa ,A new solution for data extraction : Gene/clone Method, In International Journal of Computer Science and Network Security, VOL.6 No.7A, July 2006

[5] Chara Aziz,Benlahmer EL Habib,Abderrahim Tragha,Mohamed Rachdi-Kamal elguemmat :Generate Tools through a Data Extraction Web Example of the production of meta social network, VOL.2 N°9, September 2012

[6] Seo, Heekyoung, Jaeyoung Yang, and Joongmin Choi. "Knowledge-based wrapper generation by using XML." IJCAI-2001 Workshop on Adaptive Text Extraction and Mining. 2001.

[7] Zhang, Xuekun, Jing An, and Wen Liu. "Research and implementation of keyword extraction algorithm based on professional background knowledge." Image and Signal Processing, BioMedical Engineering and Informatics (CISP-BMEI), 2017 10th International Congress on. IEEE, 2017.

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

[8] Burke, Eric, Alan Wada, and Brian Coe. "Electronic document information extraction." U.S. Patent No. 9,547,648. 17 Jan. 2017.

[9] Kim, Hongbin, Junegak Joung, and Kwangsoo Kim. "Semi-automatic extraction of technological causality from patents." Computers & Industrial Engineering 115 (2018): 532-542.

[10] Manning, Christopher, et al. "The Stanford CoreNLP natural language processing toolkit." Proceedings of 52nd annual meeting of the association for computational linguistics: system demonstrations. 2014.

[11] Laddha, Shilpa S., and M. Jawandhiya Pradip. "An Exploratory Study of Keyword based Search Results." Indian J. Sci. Res 14.2 (2017): 39-45.

[12] Tairas, R., & Gray, J. (2009). An information retrieval process to aid in the analysis of code clones. Empirical Software Engineering, 14(1), 33-56.

[13]Zoeter, O., Taylor, M. J., Snelson, E. L., Guiver, J. P., Craswell, N., & Szummer, M. (2011). U.S. Patent No. 8,037,043. Washington, DC: U.S. Patent and Trademark Office.

[14] Wang Y. ,Semantic information extraction for software requirements using semantic role labeling

(2016) Proceedings of 2015 IEEE International Conference on Progress in Informatics and Computing, PIC 2015, , art. no. 7489864 , pp. 332-337.

[15] Yadav, M. P., Feeroz, M., & Yadav, V. K. (2012, July). Mining the customer behavior using web usage mining in e-commerce. In Computing Communication & Networking Technologies (ICCCNT), 2012 Third International Conference on (pp. 1-5). IEEE.

[16] Kara, S., Alan, Ö., Sabuncu, O., Akpınar, S., Cicekli, N. K., & Alpaslan, F. N. (2012). An ontology-based retrieval system using semantic indexing. Information Systems, 37(4), 294-305.

[17] Bahafid, A., El Guemmat, K., & Talea, M. (2015). Toward a new Information retrieval system based on an e-commerce ontology. International Journal of Advanced Studies in Computers, Science and Engineering, 4(10), 26.

[18] Arp, Robert, Barry Smith, and Andrew D. Spear. Building ontologies with basic formal ontology. Mit Press, 2015.

[19] Hepp, M. (2008, September). Goodrelations: An ontology for describing products and services offers on the web. In International Conference on Knowledge Engineering and Knowledge Management (pp. 329-346). Springer, Berlin, Heidelberg.

[20] TAO, Teng-yang, and Zhao Ming. "An ontology-based information retrieval model for vegetables e-commerce." Journal of Integrative Agriculture 11.5 (2012): 800-807.