Technological Tools Integration and Ontologies for Knowledge Extraction from Unstructured Sources: A Case of Study for Marketing in Agri-Food Sector

Adriana Caione, Salento University, Engineering Innovation Department, Lecce, Italy,
adriana.caione@unisalento.it
Roberto Paiano, Salento University, Engineering Innovation Department, Lecce, Italy,
roberto.paiano@unisalento.it
Anna Lisa Guido, Salento University, Engineering Innovation Department, Lecce, Italy,
annalisa.guido@unisalento.it
Monica Fait, Salento University, Economic Sciences Department, Lecce, Italy,
monica.fait@unisalento.it
Paola Scorrano, Salento University, Economic Sciences Department, Lecce, Italy,
paola.scorrano@unisalento.it

Abstract

With the advent of Internet and Social networks, the process of knowledge extraction and search has became of great importance, especially for companies that are interested in collecting customer feedback, establishing a brand presence, observing the way their brands are discussed and perceived. Several tools for the extraction of semantically related concepts from unstructured sources are available on the web. These sources such as forums, blogs, etc. can provide real-time information about how users perceive certain brands and can guide companies in making market decisions. In this paper, we propose an innovative knowledge extraction architecture realized through the integration of some existing java based tools that aims to retrieve more frequent concepts from unstructured sources, suggest other links of articles and images, and detect the language used in the sources. The architecture provides a knowledge base of a specific domain, which is used to suggest other concepts related to the research and to filter the results obtained from the elaboration of the unstructured sources. This software, integrated with traditional marketing channels (analysis of reports) makes possible the extraction of useful information for companies. In order to show, interpret and benefit from the software output we present a case of study related to marketing in agri-food sector.

***Keywords**:* Knowledge extraction, unstructured sources, marketing intelligence, agri-food sector.

Introduction

In recent years, web and information technologies development has led to a dramatic increase of the information available in electronic format. Only a small part of them is contained in digital sources in a structured form, while the majority is in a semi-structured or unstructured form.

To extract information and then knowledge from structured sources, consolidated handling languages and powerful interrogation techniques are available.

Unlike structured data, unstructured data do not have an identifiable structure. Examples are images, videos, emails, documents and texts. The data in web pages represented in a mark-up language such as HTML, are considered unstructured (ITL Education Solutions Limited, 2008). They are human readable, but of difficult tractability using software tools. Complexity depends on the intrinsic ambiguity of natural language. As part of companies and institutions, the use of techniques, that can extract information of interest from the moles of textual materials that must be managed daily, has thus become a priority.

In Knowledge Discovery, specific techniques of Text Mining are necessary to extract information from such unstructured data (Rajman and Besançon, 1998).

The Information Extraction (IE) is a technique for the extraction of information with the purpose of acquiring relevant concepts from unstructured sources (Wu, 2002). It exploits also techniques for Natural Language Processing (NLP), a process for the automatic analysis of information in written or spoken natural language. It is a semantic search that, trying to get closer to the mechanism of human learning, returns results containing semantically related concepts to each other.

Often these tools and techniques are supported by ontology that is a formal, explicit specification of a shared conceptualization (Studer, Benjamins and Fensel, 1998).

The semantics and ontologies maximize the value of the process of information and document management, automate and simplify the processes of analysis and classification of information and documents according to taxonomic rules constantly updated.

Different tools of Knowledge Extraction and Discovery have been developed. For example in a study by Gangemi (2013) in which many tools are described and compared by defining and using some measure parameters such as performance, functionality, etc. The analysis and tests show the importance that may result from the integration of measurement and functionalities of different tasks in the perspective of providing useful analytic data out of text.

An attempt to combine diverse NLP algorithms with a variety of strengths and weakness is realized in the framework FOX (Federated knOwledge eXtraction Framework, *http://aksw.org/Projects/FOX.html*). It is discussed and used in a study by Ngomo, Heino, Lyko, Speck and Kaltenböck (2011), which talks about the importance of integrating frameworks like this into CMS (Content Management System) in order to allow knowledge extraction from unstructured data.

Besides tools/algorithms integration, another aspect of great importance is the multi-language support. It is argued in (Gerber and Ngomo, 2012) in order to make the approach of knowledge extraction independent of the language in which the text is written.

This paper makes a combined use of information extraction tools from unstructured sources, mostly from blogs, forums and social networks, and of ontologies that describe the specific domain, in order to extract information related to those sought in the documents analysed.

In this paper we explain the use of different features, both back end (concepts extraction, multi-language, etc.) and front end (like the tag cloud visualization), offered by the various existing tools and the use of ontologies to provide a tool of knowledge extraction which allows, through the use of ontologies, information extraction from different unstructured sources on the base of user-defined keywords. The output can be a useful analysis tool for marketing experts who want to explore web sources in order to identify, for instance, new markets. The multi-language aspect facilitates the search on more foreign countries.

Once a context and a reference domain have been defined, it is important to identify the input variables, such as unstructured sources and keywords as more specific to the sector, and identify or define an ontology that describes the reference domain through concepts and relationships.

In Section 2, we present the knowledge extraction tools used in combined manner within the proposed architecture. In Section 3, we describe the architecture, focusing on the integration of the tools and the ontologies, and the software operation flow. In Section 4, we propose a case study relative to the agri-food sector in order to show how the system works in a specific domain.

Knowledge extraction in marketing

The increasingly intense competition between companies, the actual period of crisis and other related factors, make the marketing an operation and a management mechanism essential for companies. To allow the market entry and get good and consistent profits, it is essential to anticipate and beat the competition. So, they need to adopt innovative and automated techniques to study customers’ needs, tastes and interests and to contribute additionally to the results obtained using the consolidated marketing tools.

The existing management systems allow an optimal management of structured information (data for which the information is readily available and interpretable by software tools); the same cannot be said for unstructured information (data contained in documents or web pages, for instance). The latter are in companies in large amount, with a growth rate greater than that of structured information, and by they can depend many business decisions.

The spread of social networks raises even more the consumer at the centre of marketing processes, imposing network content monitoring, listening to the conversations and creating "conversational relationships" based on the application of technical and marketing tools that exploit the potentiality of a bi-directional communication path.

Particular relevance assumes the ability to understand, highlight and extract unstructured most relevant content (Sentiment Analysis) in order to transfer value to business processes. The exploitation of unstructured information allows the monitoring of the target market (market sensing) and the acquisition of data useful for designing marketing activities (market insight) (Kotler, Keller, Ancarani, and Costabile, 2012).

Semantic Web Technology

Web technology has evolved from Web 1.0, to Web 2.0, to Web 3.0, starting from Tim Berners-Lee's inventing the World Wide Web in 1989. Web 1.0 is primarily a one-way publishing medium and information-centric (Murugesan, 2009). Web 2.0 links people and users, and with user generates content capability (Murugesan, 2009), it is people-centric. Web 3.0 is the Semantic Web, which is defined as a mesh of information linked up to be easily processable by machines, on a global scale (Siau and Tian, 2004).

The introduction of ontologies comes as part of the semantic web, from the need to have a language for domain representation that allows expressing the meaning of the documents present in the network. Therefore, the main goal is to structure the knowledge that is, answering to the three questions: research, extraction and maintenance of information.

The peculiarity of the ontology is the use of common words and concepts to describe and represent the domain of interest; this makes it understandable and usable by people, applications, databases, etc. in order to share a common knowledge concerning any domain.

There is no single correct ontology-design methodology (Noy and McGuinness, 2001) many of them are discussed in (Fernández López, 1999) which presents and compares the most representative methodologies.

A widely used methodology in ontology development is described in (Noy and McGuinness, 2001), whose steps are:

1. Determine the domain and scope of the ontology;
2. Consider reusing existing ontologies;
3. Enumerate important terms in the ontology, those on which users want to make statements;
4. Define the classes and the class hierarchy. It is necessary to specify classes and organize them into a taxonomic hierarchy (i.e. subclasses and superclasses);
5. Define the properties of classes. They can be of different types: "intrinsic" (such as the taste of a wine), "extrinsic" (as the name of a wine), constituent parts, in the case of structured objects (for example parts of the body), or relationships to other entities;
6. Define the property details. Each property can have several characteristics; cardinality, value type, domain and range;
7. Create instance.

Knowledge Extraction Tools

The development of semantic marketing intelligence software was preceded by a technology scouting, in order to know the state of the art and the tools currently available for the extraction of concepts from structured and unstructured sources.
As it regards the tools, we analysed:

- AlchemyAPI (*http://www.alchemyapi.com*): it is useful to extract the main ideas contained within a web page (source unstructured) and to retrieve authors;
- Zemanta (*http://developer.zemanta.com*): it is used to retrieve images and articles related to the source;
- Jsoup (*http://jsoup.org*): useful to extract the content of a page, with no html tags, starting from the provided address;
- WordReference (*http://www.wordreference.com/docs/api.aspx*): it is used to make translations of the keywords in the languages of the sources.

AlchemyAPI

AlchemyAPI technology uses natural language processing and machine learning algorithms to extract semantic metadata from a text, such as, information about people, places, companies, topics, facts, relationships, authors and languages.

The API endpoints are oriented to performing content analysis of web pages accessible from the Internet, html pages or textual content.

Among the features available there is the possibility of extraction of entities, concepts, text categorization, extraction of relations, language detection, extraction of words, sentiment analysis, text extraction, etc., We analyse those which it is made use within the semantic marketing intelligence software.

Language Detection

AlchemyAPI provides functionality for the recognition of the language of a text, HTML page or web-based content. It identifies more languages than other services of text analysis, with extremely high accuracy rates.

Keyword Extraction

AlchemyAPI is able to extract keywords from a textual, HTML or web-based content. Statistical algorithms and natural language processing technologies are used to analyse data, to extract keywords that can be used to index contents, to generate tag clouds, etc.

This processing is supported in different languages.

Author Extraction

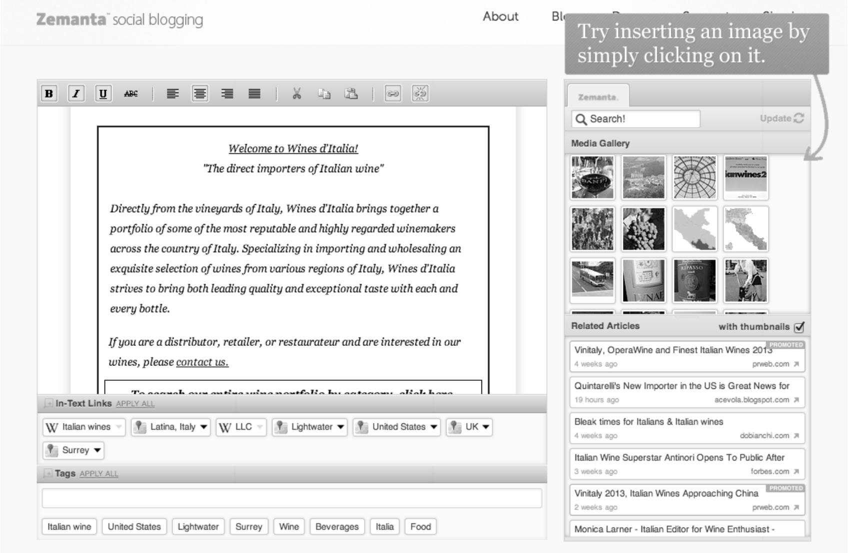
AlchemyAPI is able to extract publisher information from web pages. If an article of news or of blog specifies an author, it attempts to extract it automatically.

Zemanta API

Zemanta is a content suggestion engine for bloggers and other creators of information.

It analyses the user-generated content (for instance a blog post), using the natural language processing and semantic search technology to suggest images, tags, and links to related articles, as in Fig 1.

It suggests content from Wikipedia, YouTube, IMDB, Amazon.com, CrunchBase, Flickr, ITIS, Musicbrainz, MyBlogLog, Myspace, NCBI, Rotten Tomatoes, Twitter, and Snooth Wikinvest, as well as Blog of other users Zemanta.



Welcome to Wines d'Italia!
"The direct importers of Italian wine"

Directly from the vineyards of Italy, Wines d'Italia brings together a portfolio of some of the most reputable and highly regarded winemakers across the country of Italy. Specializing in importing and wholesaling an exquisite selection of wines from various regions of Italy, Wines d'Italia strives to bring both leading quality and exceptional taste with each and every bottle.

If you are a distributor, retailer, or restaurateur and are interested in our wines, please contact us.

Vinitaly, OperaWine and Finest Italian Wines 2013 PROMOTED
4 weeks ago prweb.com

Quintarelli's New Importer in the US is Great News for
19 hours ago acevola.blogspot.com

Bleak times for Italians & Italian wines
4 weeks ago dobianchi.com

Italian Wine Superstar Antinori Opens To Public After
3 weeks ago forbes.com

Vinitaly 2013, Italian Wines Approaching China PROMOTED
2 weeks ago prweb.com

Monica Larner - Italian Editor for Wine Enthusiast -

Fig 1. Suggested contents by Zemanta©

Zemanta is a service that connects well-known databases in a single-point solution to detect other content.

The software uses the extraction feature of articles and images.

Jsoup API

It is a Java library for HTML content processing. Provides an API for extracting and manipulating data. Using DOM, CSS, jQuery-like methods, parses the HTML from a URL, file, or a string, finds and extracts data, uses DOM and CSS selectors, manipulates HTML elements, attributes, and text.

Proposed Architecture

Based on the above considerations, in this section, we describe the proposed architecture that integrates existing java based tools and ontologies in order to retrieve interesting content for the specific domain.

The core of the marketing intelligence system is composed of the following modules (Fig 2.):

1. *Information Extractor*: searches keywords within reference ontology in order to identify other semantically related words. Makes a research of the above words in the sources and calculates the occurrences. Uses the sources and keywords to search for other sources linked to the first. It is possible to edit/enrich the knowledge model;
2. *Knowledge Filter*: filters the extracted information based on the model of knowledge;
3. *Knowledge Presentation*: displays this information to the user in a consistent manner. It is possible to enrich/update these sources at any time.

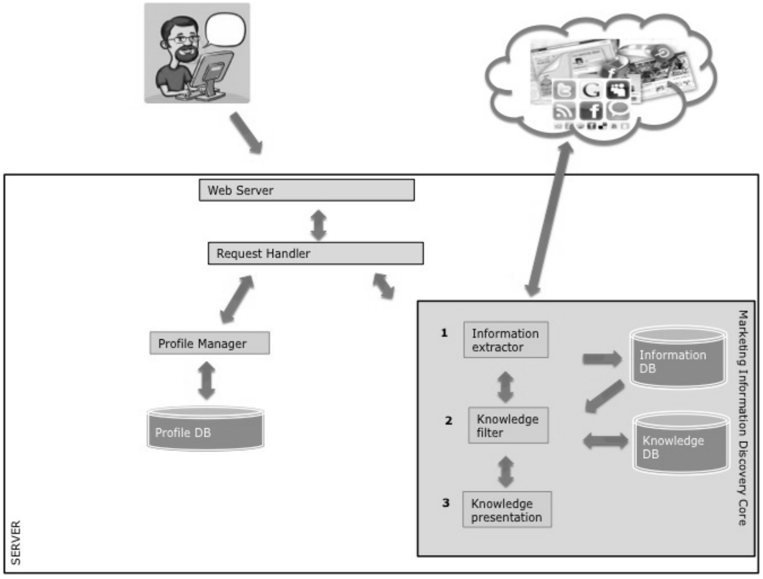


Fig 2. Reference architecture

In addition to the knowledge model, the *Profile DB* contains information about registered users, the *Information DB* stores the sources, the keywords associated with the sources and the results of the researches carried out.

This architecture is behind the system, which analyses unstructured sources by:

- Allowing, for a well defined research context, classification and analysis of the occurrences of specific information (knowledge gathering);
- Providing a single entry point to information resources on a specific domain in support of the user;
- Proposing to the user a set of information to be used for its analysis.

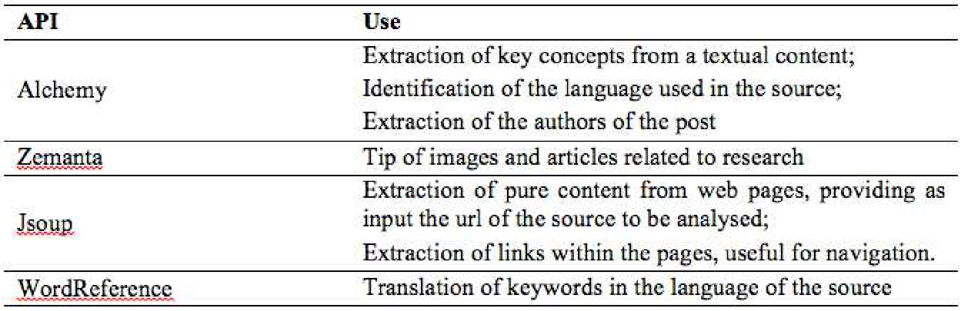
We now describe each module that constitutes the architecture here introduced.

- **Request Handler:** it receives the requests of the user. It accesses to the profile database in order to retrieve information of the authenticated user, or to store data related to the user. On the other side, it forwards the requests to the module of marketing information discovery. Both this module of information discovery and the profile manager answer to the Request Handler with information of the authenticated user or related to him, output of the elaboration in the form of a tag cloud of relevant concepts or of a table containing articles or images related to the research. This module is well suited to be used with requests (sources and keywords) in different languages. The software, using AlchemyAPI and WordReference, detects the language of the sources and keywords, and if there are differences, it translates the latters in the language of the sources. By this way, users can carry out investigations researching sources related to countries and languages different from their own and get useful results for its research.
- **Profile Manager:** this module is an intermediate level between Request Handler and Profile DB. It makes queries to the database to retrieve information related to the authenticated (personal data, relation to personal sources, relations to researches and output of the resourced processing), and to store other ones.
- **Profile DB:** it stores information about the users that are interesting in using the marketing intelligence system. Not only personal information, therefore, but also it keeps track of the relations to transactions carried and to the sources related to the user.
- **Information Extractor:** this module integrates the existing java based tools, Jsoup, AlchemyAPI and WordReference, with the aim of analysing sources and keywords selected/entered by the user in the system. The inputs can be provided in different languages. At the base there is a mechanism of language detection (through AlchemyAPI) used within each source and each keyword, and a subsequent translation (through WordReference) of the keywords in the language of the source.

- **Knowledge filter:** it makes use of ontology (one or more) of a specific domain in order to extract other words or concepts related to the research that will be processed by the Information Extractor module.
- **Knowledge presentation:** at the end of the sources and keywords processing, this module prepares the output to be presented to the user in the form of tag cloud or table of articles and images (extracted by Zemanta API), research results, etc.
- **Information DB:** it stores information sources from which to extract the data and the keywords to search, specific to the agri-food field that user want to inspect. These sources can be stored in shared manner or linked to the profile of a user and shared later. The database also contains all of the research performed by the users and the results of the processing, in order to have an historical readily available.
- **Knowledge DB:** is the ontology of the specific agri-food field. It contains the description of the domain through concept semantically related. The ontology is used in two direction:
    - Before processing sources and keywords, in order to suggest other words related to the research;
    - After processing sources and keywords, to filter the output deleting the words not related to the specific domain.

Tools Integration

The present paper properly integrates the APIs discussed in Chapter 2 for the realization of semantic marketing intelligence software. Fig 3. summarizes the features used for each API.



API	Use
Alchemy	Extraction of key concepts from a textual content; Identification of the language used in the source; Extraction of the authors of the post
Zemanta	Tip of images and articles related to research
Jsoup	Extraction of pure content from web pages, providing as input the url of the source to be analysed; Extraction of links within the pages, useful for navigation.
WordReference	Translation of keywords in the language of the source

Fig 3. Features used for each API

Semantic Ontology Integration

We adopt an ontology model specific to the sector.

An ontology, for example relative to the wine sector, was obtained on the web (*www.w3.org/TR/owl-guide/wine.rdf*), and reused. It can be integrated with other information over time and it is possible to replace or complement it to other ontologies for specific sectors.

The ontology downloaded from the web and reused for example, contains descriptions of hierarchies and categories of foods and wines, along with restrictions concerning the association of particular instances.

Wine

The ontology describes the concept of wine. According to the specifications, a wine is a drinkable liquid product from at least one winery, composed of at least one type of grape. Wine has four properties: colour, sugar, body and flavour.

Meal Course

The concept of meal course emphasizes the combination of a meal with a wine. Each meal course includes at least one food and one drink, the latter is expected to be a wine. When the user selects a meal course, or a single food connected to a meal course, it is possible to have suggestions on wine to approach or vice versa. The ontology is used before the sources and keywords processing by the system with the aim to extract semantically related words and concepts to those specified by the user.

It may also be used after elaboration with the purpose of filtering the results obtained.

Details of the Integration

In this section we analyse the integration of existing java based tools, ontologies and databases from an implementation point of view.

Using the API provided by the library Jsoup, it is possible to extract pure content from a web page, given the address of the same. HTML tags related to the remaining links, media and import are then deleted from the content.

Besides extracting the content of the input sources, are extracted links (of number equal to that indicated by the user) contained in the page. In order for the link whose contents are consistent and relevant to the source, we extract links that allow navigation of the website (it checks that the protocol and the host of the URL are the same of the source) and do not relate to such contacts, categories, policy, about, etc.

Just as the first ones, these sources are processed through Jsoup to extract only the content.

Before proceeding with the search of the keywords in the selected and extracted sources, the system checks whether it is possible to combine words entered with other semantically related. Using SPARQL query and the framework Jena the system asks sector-specific ontology to extract other concepts connected to the first ones.

The content previously extracted from the source is passed as input to the method provided by AlchemyAPI that detects the language used. It also identifies the language of keywords. This information is used with WordReference API to translate keywords in the language of each source.

For each source, the system proceeds with the extraction of concepts and most recurrent words (AlchemyAPI). The number of occurrences in the text of each concept is calculated.

In addition, the system checks the presence or absence of the keywords selected/entered by the user. If the API has already extracted them, the value of its occurrences is increased; otherwise, they are added to the list of words extracted. This list is used to create a tag cloud visible to front end.

The research results can be filtered on the basis of the ontology. Given the keywords, the system extracts the concepts related to them and the words that are not in this set are excluded from the result.

Zemanta API suggests contents such as images and articles. This suggestion is done on the web in real time.

The results obtained are stored in the database in correspondence with the profile of the user logged. In this way, it keeps a history of researches always available and viewable instantaneously, without the need of having to perform elaborations.

System Operation Flow

The flow of operation of the software is characterized by an initial phase of identification of sources (unstructured) from which the user can extract information. This allows the creation of a database that can be customized through the classification by type of sources (blogs, websites of buyers, websites of restaurants, etc.) and geographical area of reference (Germany, USA, Australia, etc.). In addition, in order to make more efficient over time the process of information extraction, there is a system of storage and subsequent handling of personal data, sources, keywords and researches. During elaboration, the software makes use of an ontology specific to the domain of interest, useful to suggest other keywords related to the research, and to filter the results of the sources.

The processing of information content begins with the selection of one or more sources and keywords, including those previously entered into the database or entered manually by the user, which is followed by an indication of the navigation depth degree within each selected source. An advanced version for advanced users, introduces an additional level of analysis, allowing user to specify for each source the most interesting sections from which to start the research.

The software can produce a tag cloud; to do this, the system (designed to support multiple languages):

1. Extracts the number of pages specified by the user in each source
2. Queries ontology to see if it is possible to combine words entered with other semantically related words
3. For each source, extracts the most recurring concepts and words and calculates the number of occurrences in the text.

The system then is able to extract images and articles related to the research carried out (depending on the sources and the previously selected keywords).

Agri-food Sector Marketing Intelligence: a Case of Study

The software testing is carried out in collaboration with the Department of Economic Sciences of the University of Salento. The experimentation concerned the sources of oil producers in the Australian market. As regards the choice of keywords, have been taken of those representing the three main categories of oil, i.e.: *olive oil*, *virgin olive oil*, *extra virgin olive oil*.

Following the system operation flow previously described, we illustrate the results obtained and how we can interpret it.

Fig 4. shows a portion of tag cloud relative to oil producers in Australia, obtained by the system elaboration, which shows the frequency of the concepts related to olive oil.

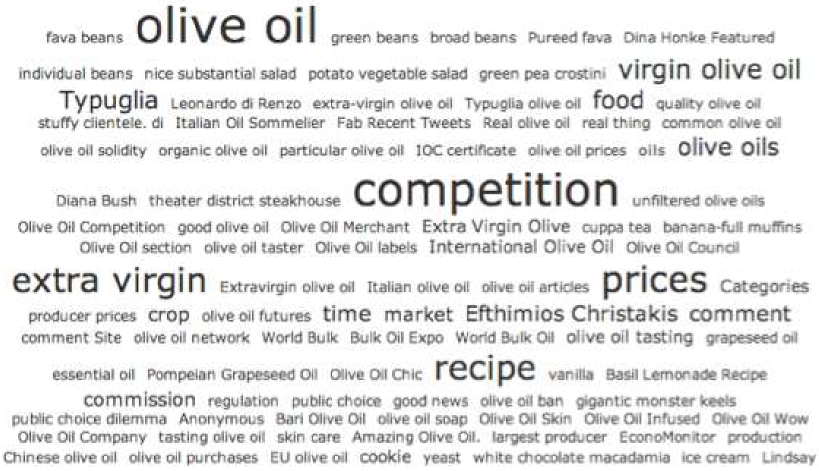


Fig 4. Tag cloud oil producers in Australia

The portions of the tag cloud of Italian and Spanish exporters of oil in Australia are respectively in Fig 5. and in Fig 6.

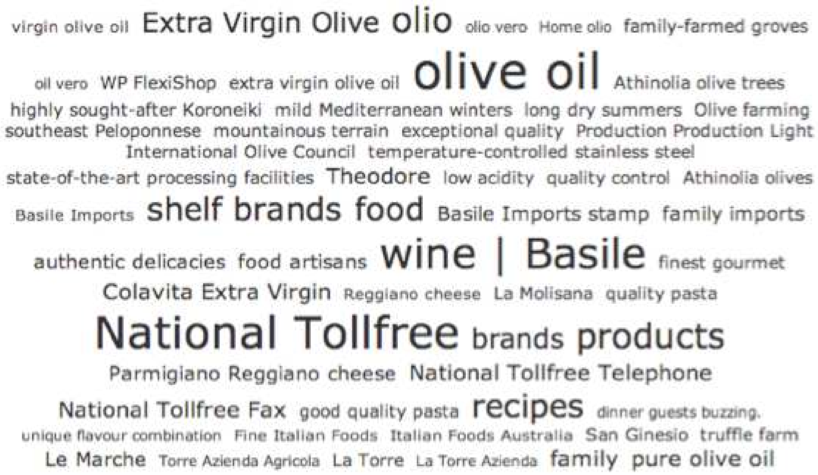

virgin olive oil Extra Virgin Olive olio olio vero Home olio family-farmed groves
oil vero WP FlexiShop extra virgin olive oil olive oil Athinolia olive trees
highly sought-after Koroneiki mild Mediterranean winters long dry summers Olive farming
southeast Peloponnese mountainous terrain exceptional quality Production Production Light
International Olive Council temperature-controlled stainless steel
state-of-the-art processing facilities Theodore low acidity quality control Athinolia olives
Basile Imports shelf brands food Basile Imports stamp family imports
authentic delicacies food artisans wine | Basile finest gourmet
Colavita Extra Virgin Reggiano cheese La Molisana quality pasta
National Tollfree brands products
Parmigiano Reggiano cheese National Tollfree Telephone
National Tollfree Fax good quality pasta recipes dinner guests buzzing.
unique flavour combination Fine Italian Foods Italian Foods Australia San Ginesio truffle farm
Le Marche Torre Azienda Agricola La Torre La Torre Azienda family pure olive oil

Fig 5. Tag cloud Italian oil exporters in Australia

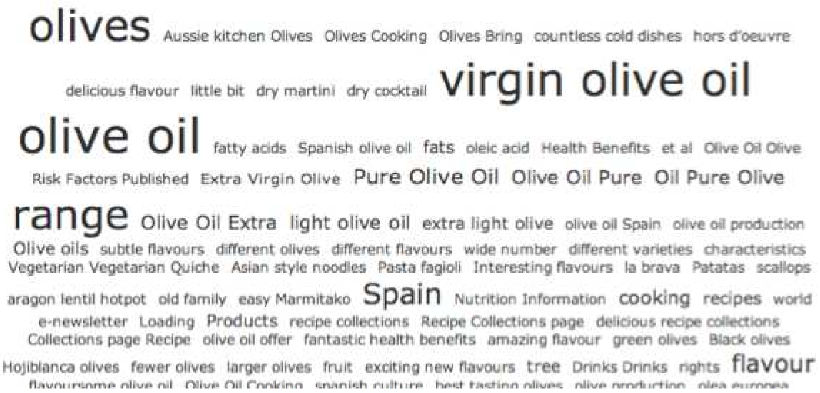

olives Aussie kitchen Olives Olives Cooking Olives Bring countless cold dishes hors d'oeuvre
delicious flavour little bit dry martini dry cocktail virgin olive oil
olive oil fatty acids Spanish olive oil fats oleic acid Health Benefits et al Olive Oil Olive
Risk Factors Published Extra Virgin Olive Pure Olive Oil Olive Oil Pure Oil Pure Olive
range Olive Oil Extra light olive oil extra light olive olive oil Spain olive oil production
Olive oils subtle flavours different olives different flavours wide number different varieties characteristics
Vegetarian Vegetarian Quiche Asian style noodles Pasta fagioli Interesting flavours la brava Patatas scallops
aragon lentil hotpot old family easy Marmitako Spain Nutrition Information cooking recipes world
e-newsletter Loading Products recipe collections Recipe Collections page delicious recipe collections
Collections page Recipe olive oil offer fantastic health benefits amazing flavour green olives Black olives
Hojiblanca olives fewer olives larger olives fruit exciting new flavours tree Drinks Drinks rights flavour
flavoursome olive oil Olive Oil Cooking spanish culture best tasting olives olive production olea europea

Fig 6. Tag cloud Spanish oil exporters in Australia

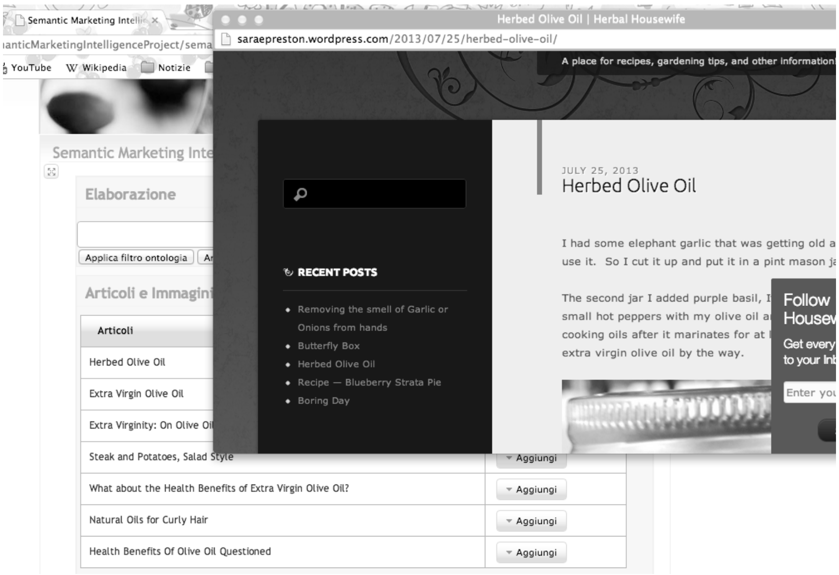
The analysis of the images can be seen as producers and exporters communicate across three groups of words: olive oil, virgin olive oil, extra virgin olive oil. The producers emphasize the very feature extra virgin, note the font size of the words Extra Virgin Olive Oil, Extra Virgin Australia in Fig 4.; exporters are more generic, the characteristic olive oil is the most present (Fig 5. and Fig 6.).

In the websites of Australian producers, there are recurrent sensory attributes that enhance the organoleptic qualities and the natural appearance of the product, such as organic, acid, natural olive oil. In the websites of exporters Italian and Spanish there are many occurrences of words that highlight the sensory attributes such as flavour and texture, fruity flavour, fluidity to name a few.

Finally, in all the sources analysed, there are concepts related to food and health benefits that may be derived from the consumption of olive oil.

The analysis shows, therefore, the willingness on the part of Australian producers to make known their product by providing different elements of national recognition, first of all to be extra virgin. On the other hand, the exporters, interested in marketing, aim at the characteristics of the sensory attributes such as flavour, acidity, and consistency, and relations with other food products.

As an alternative or as a complement to the tag cloud, user can view other interesting information such as articles related to the research (Fig 7.).


Semantic Marketing Intelli
anticMarketingIntelligenceProject/sema
YouTube
Wikipedia
Notizie
Semantic Marketing Inte
Elaborazione
Applica filtro ontologia
Articoli e Immagini
Herbed Olive Oil | Herbal Housewife
saraepreston.wordpress.com/2013/07/25/herbed-olive-oil/
A place for recipes, gardening tips, and other information!
RECENT POSTS
Follow
Housew
Get every
to your Inb
Enter you

Articoli	
Herbed Olive Oil	
Extra Virgin Olive Oil	
Extra Virginity: On Olive Oi	
Steak and Potatoes, Salad Style	Aggiungi
What about the Health Benefits of Extra Virgin Olive Oil?	Aggiungi
Natural Oils for Curly Hair	Aggiungi
Health Benefits Of Olive Oil Questioned	Aggiungi

- Removing the smell of Garlic or Onions from hands
- Butterfly Box
- Herbed Olive Oil
- Recipe — Blueberry Strata Pie
- Boring Day

JULY 25, 2013

Herbed Olive Oil

I had some elephant garlic that was getting old a use it. So I cut it up and put it in a pint mason ja

The second jar I added purple basil, I small hot peppers with my olive oil a cooking oils after it marinates for at l extra virgin olive oil by the way.

Fig 7. Related articles

Managerial Implication

Systematizing the extracted information it was possible to acquire:

- Information on the reference market (*market sensing*), the demand of olive oil is not satisfied only by traditional exporting countries such as Italy, Spain and Greece, but also by local producers entering the market. The website analysis has allowed us to see how Australian producers share a consistent communication strategy whose goal is to enhance the national product through the communication of three basic elements: sensory and quality elements of the product, the recourse to modern techniques of cultivation and production of olive oil environmentally friendly, the food and the health benefits. In contrast, the communication of Italian and Spanish producers in the Australian market continues to focus only on the traditional sensory dimensions of the product, probably because often it passes through buyers;
- Quantitative data, useful for designing marketing actions (*market insight*). Content analysis of blogs and forums on olive oil has allowed us to observe that the sensory and gustatory elements, together with experiential items (*food and health benefits*), are the main drivers of recognition of the three types of product *olive oil*, *virgin olive oil*, *extra virgin olive oil*; rarely there is a direct association to specific brand, more frequent however, is the reference to the origin country.

Conclusion and Future Works

In this paper we show an architecture that integrates existing java based tools for information extraction, and ontologies in order to extract relevant concept starting from a set of unstructured sources of a specific domain and a set of keywords of interest and to generate output easily interpretable by the users of the system. The integration is useful because each tool has some strengths and weakness as described in the study by Gangemi (2013) or has different characteristics. The architecture and the software modules described in this paper make use of the features of the tools that better ensure the achievement of the results and, unlike the study by Ngomo, Heino, Lyko, Speck and Kaltenböck (2011), they are calibrated to be integrated into a web application usable by end-users (for example, the marketing manager) from and for different countries since the system is independent from the language of the sources analysed.

Managerial implications are object of study of marketing managers of involved companies to define future strategies. The instrument will see the integration of a section focused on the analysis of structured sources and of useful guidelines for interpreting the results of the performed analyses.

References

Fernández López, M. (1999), 'Overview of Methodologies for Building Ontologies' Proceedings of the International Joint Conference on Artificial Intelligence (IJCAI) Workshop on Ontologies and Problem Solving Methods KRR5, 31 July - August 6 1999, Stockholm, Sweden.

Gangemi, A. (2013), 'A Comparison of Knowledge Extraction Tools for the Semantic Web' Proceedings of European Semantic Web Symposium (ESWC), Lecture Notes in Computer Science (LNCS), Springer, 26-30 May 2013, Montpellier, France.

Gerber, D. and Ngomo, A.-C. N. (2012), 'Extracting Multilingual Natural-Language Patterns for RDF Predicates' Knowledge Engineering and Knowledge Management (EKAW), Lecture Notes in Computer Science, Springer, 8-12 October 2012, Galway City, Ireland.

Kotler, P., Keller, K. L., Ancarani, F. and Costabile, M. (2012), 'Marketing Management' Pearson Education.

Murugesan, S. (2009) Handbook of Research on Web 3.0, 2.0, and X.0: Technologies, Business and Social Applications, Information Science, Murugesan, S. (Editor), Chapter 1, 1-11.

Ngomo, A.-C. N., Heino, N., Lyko, K., Speck, R., and Kaltenbock, M. (2011), 'Scms-semantifying content management systems' International Semantic Web Conference (ISMC), Springer, 23-27 October 2011, Bonn, Germany, 189-204.

Noy, N. F. and McGuinness, D. L. (2001), 'Ontology Development 101: A Guide to Creating Your First Ontology' Technical report, KSL-01-05, Stanford Knowledge Systems Laboratory.

ITL Education Solutions Limited (2008) Introduction to Database Systems, Pearson Education India Chapter 17.

Rajman, M. and Besançon, R. (1998), 'Text Mining - Knowledge extraction from unstructured textual data' 6th Conference of International Federation of Classification Societies (IFCS), 21-24 July, Rome, Italy.

Siau, K and Tian, Y. (2004), 'Supply Chains Integration: Architecture and Enabling Technologies' *The Journal of Computer Information Systems*, 44(3), 67-72.

Studer, R., Benjamins, R. and Fensel, D. (1998), 'Knowledge engineering: Principles and methods' *Data & Knowledge Engineering*.

Wu, T. (2002), 'Theory and Applications in Information Extraction from unstructured text' Theses and Dissertations, Paper 741.