THE INDEXING AT THE INTERNET

Isidoro Gil-Leiva
Universidad de Murcia
Murcia/Spain

ABSTRACT

It is presented an analysis of the presence of indexing at the Internet supported by the idea that good part of the basic pillars in the ones which this web of computers is based is impregnated in higher or lower measure by indexing. The pillars are the metadata, the searchers, the users and the web positioning. Each one of these elements is reviewed to prove the approach which is accomplished the one is denominated **Universe of the Web Indexing.**

Keywords: Indexing; Internet; Metadata; Web Positioning; Users; Searchers; Search Engines; Universe of Web Indexing.

INTRODUCTION

Before the Internet appearance, most of the occasions, the data, the information and the steps were distant and decentralized. To accomplish an administrative step it was necessary to go the administration, to read or buy a book to go to the library or to a bookstore, to contemplate determined museum work go to one, or to buy a car move to the car dealer and, like that, successively.

Along with the Internet generalization almost all the activity branches of the human being were blooming on the web and, consequently, the competence of the physical world in the services, trade or culture were translated to the digital scope. However, according to the improvements of the content at the web it was necessary the development of bridges, to connect the people to the information mass available at the web, with the purpose to access in a fast and efficient way this new environment. These footbridges are the denominated search or research engines.

At this new reality it was created a space which we name *Universe of Web Indexing* or *Environment of Web Indexing* which is composed by four distinct scopes, but tremendously interrelated. On one hand, the marking languages and encoding standards which facilitate the organization and dissemination of the information by

47

the web; on the other hand, the Web owners (companies, institutions or privates) that wish their contents to have the highest visibility, what means, a good position at the researches, for dealing with service improvements, prestige or profitability, among other variables; at the third space, are the search engines which use complex algorithms to offer a ranking of the information found to satisfy to the clients; and the fourth scope, the users of the search engine, also, employ tactics to maximize the effort and time employed in the use of engines.

These four which is constituted at the Indexing Universe is impregnated by the indexing, and got to this *Indexing Universe* by the progressive extension, not only of knowledge and practice of the indexers but also the information and documentation of professionals in general, with a view to the popularization of the Internet. The generalization of these concepts and practices is a reality and easily identifiable in numerous examples, as the one appeared in a Spanish news of general character and national circulation, in which five columns and with a extension of half page it was read[1]:

> […] Our language involves many countries, but at the Web it is constituted a unique space accessible to the researchers. The Google, for example, does not index all the Web pages in Spanish […] No researcher meets all the pages in one language: index only the most important part of its Web […] The second requirement is that the pages facilitate the work of search engines, making it possible to the contents be accessible without barriers to any researcher (instead of drowning its data in animations). Like that, today it will be possible to index them at the Yahoo or Google and, tomorrow, who knows?

At this same news, some months later, in another section, it was read:

> To run this work they were used two processes and methodologies based in indexing, relevance and popularity. The indexing is the set of actions to accomplish on the web page, for that the researcher may access all the information which it contains.

At this article of the same press, without going farther, it is talking about the four *Indexing Universe* which we mentioned previously (indexing concept, web positioning, researchers and metadata).
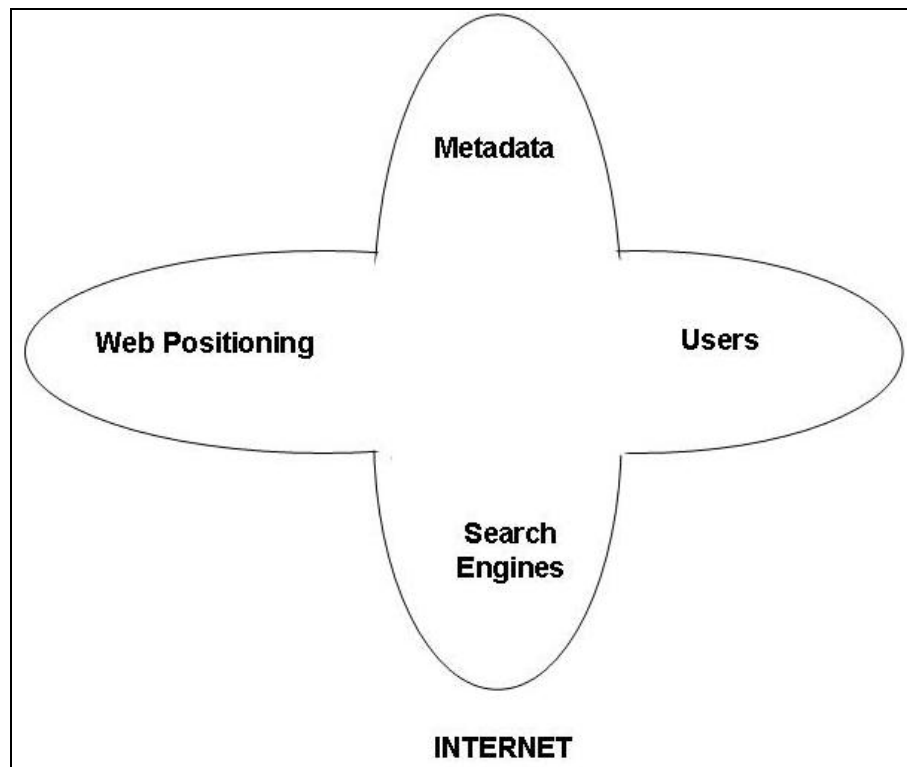
Picture 1: Web Indexing Universe.
Source: Prepared by the author.

2 INDEXING UNIVERSE AT THE INTERNET

To continue, we will hold in each one of the components of the Web Indexing Universe which we have just mentioned.

2.1 Metadata

The metadata are intended to order and describe the information contained in a document understood as object, in a way that emerge as revealing, both of the formal description and the content analysis, aiming to improve the access to these objects of information at the web. They are not more than structures of organization of information, readable by machine, whose purpose it to make the data useful, in different ways, according to the concrete necessities of each service of digital information and according to the application which are granted to them[2].

There is a wide catalog of typologies of metadata proposed by many authors, but that adjust itself perfectly to our interests is the following:

> Metadata, independent of content which gathers information about the location of a document, its creation date, change etc […] Metadata, dependent of the content, which agglutinate data about the representation and structure of an amount of information which they describe. On its turn, these metadata divide themselves in metadata based in direct content, such, as for example, the indices of a document in complete text or the color and the shape of a digital image, and descriptive metadata, such as, descriptors and identifiers, what means, the metadata which contain the description of a document without using expressly its content[3].

The sets of metadata impregnate brands or tags which are ideal peers of words or acronyms with a high value semantic and nemotechnical circulated by angles, between the ones it is located the structured information. The first tag indicates that there starts an amount of determined information and the second tag which has an oblique slash, indicates the end. These tags are easily readable, both by informatic programs and by humans and have the facility to present, structure and interchange information between computers.

```
<name> Antonio Gil Cuenca </name>
<birth place> Águilas </birth place>
<city> Murcia </city>
<address> Calle San Vicence, 7 </address>
<kinship> grandfather </kinship>
```

From the set of rules Standard Generalized Markup Language, better known as SGML and converted into ISO, in 1986, appeared the marking languages which serve to code a document through a set of tags. After the SGML appeared the language eXtensible Markup Language (XML), which is wider and converted almost into a standard. After these models proliferate marking languages, the most to specific scopes, which serve both to schematize and distribute information of any type (language HTML or XML, for example) and for disciplines or specific areas (EAD and EAC to the archive; CMI to Museums; MPEG-& to multimedia content: OWL to share ontologies at the web; ID3 to audio files MP3; MDL to audiovisual content: MCM to medicine etc). To these marking languages, there is to add other encoding

standard created in any case previously, but that share the interchange philosophy as, for example, MARC, ISAD(g) or MoReq.

We will read over again now, some marking languages and encoding standards, emphasizing the metadata which may cover the indexing.

2.1.1 HTML

The marking language Hyper Text Markup Language (HTML) uses the section of heading to transmit to the web server the information about the document. All the information which is provided at the heading is comprehended between the tag <head> and the tag </head>. Exist a series of tags reserved specifically to the heading as the one of <title> and <title> or the tag META keywords, which serves to inscribe there key-word or meaningful phrases (to the indexing) and, like that, to the search engines to exact content of the web page (to the recovering).

Source Code of HTML with the Tags META of Keywords
<head>   <meta name="generator" content="HTML Tidy, see www.w3.org" />   <title>UKOLN</title>   **<meta name="keywords" content=** "national centre, digital information management, cultural heritage, library, awareness, research, information services, public library networking, bibliographic management, distributed systems, metadata, resource discovery, conferences, lectures, workshops" **/>**   …   …

The tag <keywords> was thought of to store relevant words or phrases aiming its recovering. Its use by the webmaster is very low and unequal to the content of some studies accomplished[4], maybe for lack of knowledge, carelessness even intentional due to the excessive abuse that was done of them in many occasions.

2.1.2 Dublin Core

A *Dublin Core Metadata Initiative* is a group of work constituted by librarians, researchers of digital libraries and information providers which started to function in Dublin (Ohio), in 1995, with the purpose de proportionate recommendations about the description of information resources and of its interchange. Dublin Core provided fifteen metadata to

The description of an informational resource. For the content (Title, subject, Source, Language, Relation and Covering), for the intellectual property (Author, Editor, Collaborator, Laws) and for the format (Data, Type, Format, Identification). The Version 1.1 of the set of metadata elements Dublin Core turned, in 2003, to be one international norm under the number ISO 15836:2003 (UNE-ISO 15836:2007)

The tags related directly to the indexing are: Subject, Keywords, Producer and Date. The *Dublin Core* defines these tags on the following way:

Name: subject
Label: **Subject and Keywords**
Definition: The topic of the content of the resource.
Comment: Typically, a Subject will be expressed as keywords, key phrases or classification codes that describe a topic of the resource. Recommended best practice is to select a value from a controlled vocabulary or formal classification scheme.

Name: creator
Label: **Creator**
Definition: An entity primarily responsible for making the content of the resource.
Comment: Examples of Creator include a person, an organization, or a service. Typically, the name of a Creator should be used to indicate the entity.

Name: date
Label: **Date**
Definition: A date of an event in the lifecycle of the resource.
Comment: Typically, Date will be associated with the creation or availability of the resource.
Recommended best practice for encoding the date value is defined in a profile of ISO 8601 and includes (among others) dates of the form YYYY-MM-DD.

Dublin Core Subject Tag for XML

```
<?xml version="1.0"?>
<metadata
<dc:title>Universitat Politécnica de Valencia</dc:title>
<dc:creator>Universidad Politecnica de Valencia</dc:creator>
<dc:subject> UNIVERSIDAD INVESTIGACION DOCENCIA CENTROS
DEPARTAMENTOS ESTUDIOS ALUMNADO PROFESORES ASIGNATURAS
```

```
PROYECTOS </dc:subject>
<dc:type> Text </dc:type>
<dc:format> text/html </dc:format>
<dc:format> 1447 bytes </dc:format>
<dc:identifier> http://www.upv.es </dc:identifier>
...
...
```

2.1.3 Encoded Archival Description (EAD)

It was initiated the work with this marking language around the decade of 90, in the *University of California, Berkeley,* to create a standard structure of data which would propitiate the interchange and access to the description instruments used by the files.

The Encoded Archival Description (EAD) is composed of three main elements: 'Heading EAD'with the tag <eadheader>, 'Preliminaries' and the tag <frontmatter> and, thirdly, the element 'File Description' with the tag <archdesc>. Out of this tag <archdesc> the other tags unfold themselves which makes it possible to represent the description instruments of the file documents.

One of the subelements <archdesc> is denominated 'Heading of Authorized Access' which has as tag <controlaccess>. Therefore, at this subelement lays the result of the grouped indexing in many tags as:

```
<subject>      subject
<corpname>   name of the institution
<persname>   name of the person
<unitdate>     date that appears in the unit which describes itself
```

2.1.4 Encoded Archival Context (EAC)

The initiative to create a set of tags to represent the archival contexts which emerged in 2001 between the Project LEAF (Linking and exploring Authority Files)

and the Program of Technologies to the Information Society of the European commission.

The Encoded Archival Context (EAC) is a set of tags which allows that all document have two obligatory elements, the heading <eachheader> and the context description <condesc>. The tag <eachheader> contains data about the control of the author description and the description context. And the context description <condesc> covers the author description. On the other side, both <easgeader> and <condesc> contain specific elements in the ones which it is sheltered the indexing result.

<identity> Identity, obligatory element which groups elements to proportionate headings of names, both authorized and alternative catalogs to identify the entity.

<corphead> Heading of corporative organ, which envolves, for example, association names, institutions, companies, expositions, expeditions, vacations, etc.

<date> Date, identify any dates which deserve encoding.

<existdate> Date of existence, to the dates of existence of the entity which is being described, as establishment and dissolution of organs and birth date or people's death.

<auth> Authority which specifies a controlled or authorized vocabulary, used to compose the authority register.

<famname> Family name, destined to a group of people close related by blood laces or people who constitute a home.

<pershead> heading of personal name, to identify a person for certain. It is part of the elements such as last name, first name, patronymic or name of origin place.

<place> place, of natural character, a politic jurisdiction such as Montes Apalaches, Baltimore, Md.; Chinatown, São Francisco; or Kew Gardens, England.

<subject> Subject, which identifies a subject associated or protected by the materials described in an instance EAD. Even like that, to this tag it is recommended the use of controlled vocabulary to facilitate the access to the subjects in and between systems of description instruments.

<persname> name of the person, to the personal names.

<corpname> corporative name, to the corporative names.

<geogname> geographic name, to the geographic names.

2.1.5 Consortium for the Interchange of Museum Information (CIMI)

It deals with a consortium of institutions and organizations to preserve disseminate the cultural inheritance and museums mainly. At this set of metadata, the indexing arranges itself in the tags listed below:

```
<date> date.
<dateRange> two dates to delimitate a period.
<geo> commom geographic name as "valley", "mountain", etc.
<geogName> geographic characteristic ("Valle de Leiva", "Monte Sinai" etc.).
<keywords> list of keywords or phrases that identify the theme or text nature.
<orgName> name of an organization.
<persName> first name of a person.
<placeName> name of a place.
<region> geopolitic region.
<country> name of a country, nation, colony, etc.
<textClass> nature or theme of a text around a classification or a catalog.
```

2.1.6 Text Encoding Initiative (TEI)

TEI is an interdisciplinar and international norm which helps out libraries, museums, editors, etc., to represent all classes of humanistic texts to the research and teaching, using for that a tag structure. Some tags thoughts to contain the indexing product are the following:

```
<keywords> List of words or phrases which identify the text theme
<person>    Individual names inside the manuscripts
<institution>  Names of institutions
<origDate>   Dates that appear inside the manuscripts
<origPlace>  Places denominated in the manuscripts
<country>    Countries
<region>     Regions
```

Once analyzed the tags used by some marking languages to contain the indexing result, we will review now known standard encoding which pursues the same purpose.

2.1.7 Machine Readable Cataloging (MARC)

The Library of the U.S.A Congress developed the format LC MARC in the decade of 1960, as a set of signalers which combine numbers, letters and symbols to add them to catalographic registers. This way, each amount of bibliographic information preceded by the signalers (such as, 300; 1#;$a; $c) may be read by the computers. Below, they are presented the tags MARC which contain epigraphies or headings of subject, of names and of dates:

Subjects	Names	Data
600 - Personal names	100 and 700 - Personal names	005
610 - Entities	110 e 710 – Institutional names	033
611 - Conferences titles		260$c
648 - Chronological terms		362
650 - Subjects		
651 - Geographic Places		
653 - Not controlled terms		
654 - Thematic faceted terms		

033 2	1951-2007
600	Coudenhove-Kalergy, Richard
600	Briand, Aristide
610 24	$aEuropean Union$x-Institututions
610 24	$aEuropean Union$x-History
650	$a Community Law
700 1	$aMiguel, Mario de

2.1.8 ISAD(g)

The International Standard Archival Description (g) was prepared by the Standard Description Standard, from the International Council on Archives, composed by archivists of different nationalities. As it is indicated in the introduction of the own standard, it is constituted in a general guide to the preparation of archival descriptions which identify and explain the content and the context of the file documents, with the goal to make them accessible and interchangeable.

The standard defines twenty six elements which divide themselves between the following seven areas: Identification, Context, Content and Structure, Access Conditions and Use, Associated Documentation, Notes and, at last, Description Control. As it is observed, none element was dedicated specifically turned to the indexing. The lack of an element turned to the indexing, obliges the professionals of file, when describing documents, to decide where to put the data relative to indexing. It is observed that some professionals put them in element *6.1 Notes.* If we observe the standard, we check that this element (6.l) "serves to consign special information or any other meaningful information not included in any other element of the description." Other institutions include information related to the indexing in the element *3.1 Reach and Content,* destinated according to the standard, to 'give a view of set (for example, time periods, geographic scope) and to accomplish a content review (for example, documental types, main subject, administrative procedures) of description, appropriated to the description level". However this element seems more adequate to include the indexing, it is not the ideal to include indexing along with the review, documental typologies or administrative procedures. A third via employed by the archivists is the inclusion of an interruption denominated *Access Points*, and that comes after the last element.

Of course, this indefinition of the standard may provoke that from one side, the elements *3.1 Reach and Content* and *6.1 Notes* may shelter very heterogeneous information. On the other side, include the indexing under *Access Points* does not seem the most adequate, because between other things, as it is known, the elements *1.2 Title, 2.1 Producer name* or *4.3 Language* are also points of access to the described document. However, it would be desirable that in the next review and the standard update ISAD(g) would include an specific element to shelter the indexing of the described document, put that one of the main objectives of this standard is the one to turn accessible the documents of File and, with no doubt, the indexing result (subjects, proper names of people, of things or places and dates or periods) it is a irreplacibe way to facilitate the researches and, finally, the access to the documents.

Despite the standard improvement, the files are making a great description work and document indexing as it is observed in the following examples. Inclusively some files are being pioneers in the use of catalogs to the indexing and the

subsequent recover of documents through the Internet, as it is the case of the *Archivo Municipal de Arganda del Rey (Madrid)* the one of the File del Reino de Valencia (Valencia).

Picture 2 – Indexing Associated to a Photograph Described Following the ISAG(g) in the Municipal Archive of Logroño.
Source: Municipal Archive Photographs. ESP AML FO n 779

1.2 Title: Grand opening of the new train station work
…
…
6.1 Note: There are white margins. The series which form them are the inventory numbers of 717 to 786

Geographic Descriptors

      Term Via:

      Place: Logroño
      Province: La rioja
      Country: Spain

Organisms and Entities:
 Origin of names: Photo Payá (photographer). Fernández Ladreda e Menéndez Valdés, José Maria (General and Minister of Public Work). González Gallarza, Eduardo (General and Air Minister). Pernas Heredia, Julio (mayor of Logroño)
Subjects: Public and official acts. Iron Train
Others: Train Station

Municipal Archive of Arganda del Rey:

Reference: ESO12800148.AMAR/DD0000200006

…

…

Title: Execution letter of Felipe III the petition of Juan Batista de Granados against Juna de Higuera by the non payment of the gains of a fish meat store and olive oil.

Content Area

    Descriptors:

    Subjects:
        Execution
        Gains
        Fish Meat
        Olive oil
        Abundance

    People:
        Bautista de Granados, Juan
        Muñoz, Sebastian
        Higuera, Juan de la

    Places:
        Valladolid (MU)

…

…

2.1.9 Moreq

In the core of European Union and in the context of data interchange between the European administrations it was made in 2001 a Model of Requirements (MoReq) to establish a Management System of Electronic Documents of File (MSEDF). At this model of requirements are destinated metadata with the goal to shelter the indexing:

> *12.4.3 Described Keywords*
> 　　The MSEDF should admit the association of terms included in a controlled vocabulary as descriptive terms referent to the subject.
>
> *12.4.22 Name Based in Keywords*
> 　　It is suitable that the denominations of expedient are based in terms included in a controlled vocabulary and in relations extracted of a catalog. Even so, it is convenient that it allows the vinculation of the calalog to the classification board.
>
> *12.7.2 Subject*

2.2 Web Positioning

The Search Engine Optimization (SEO), means, 'Optimization of Search Engine' is a set of techniques applied to a web page, aiming to obtain a better position in the lists offered by the search engines, when a determined research is accomplished. At this sense, it is pointed out that it is used more generally as 'web positioning'. Since its appearance around 1990, it was developed a specific market to work with this, not only companies and professionals SEO but also informatic programs.

The Web Positioning soon emerged a great interest translated in numerous press literature, but, above all, abundance in the proper Internet[5].

To get a good Web Positioning, however it is not clear that it is possible, because each search motor uses different criteria when offering results, it is necessary a numerous set of techniques. These tactics are known as "factors SEO" and they group both the endogenous factors which perform inside the web page (optimization to the title, the content, the tags, etc) and the exogenous factors (Page rank, text at the links, external links etc). In table 1 they are shown many classifications of factor SEO ordered by relevance[6].

Table 1- Factors SEO Ordered by Relevance

Top 10	Top 15
1. Tag title 2. Text at the links 3. Use of keywords in the document 4. The document access 5. Internal links 6. The main subject of the web page 7. Links to external web pages 8. Popularity at the specific community 9. General popularity 10. Excessive repetition of keywords	1. Tag title 2. Keywords used in the document 3. Structure of Internal links 4. Unique content 5. Links to external web pages 6. Web site antiquity 7. The meta tag Description 8. Keywords at the URL 9. Tags title and Alt 10. Tags boldface, underlined, H1 11. Page density 12. Page language 13. Content update 14. Page size 15. Validation W3C

It is presented (annex 1) a compilation of many recommendations to obtain a good web positioning. In a brief analysis of the SEO factors it is observed the presence of elements very close to the theory and the indexing practice.

2.3 Searchers

The Information recover at the Internet is possible through the searchers. These may be classified in thematic indices or directories and in search engines. The Table 2 contains a comparison of the directories and search engines.

Table 2 – Comparison Directories versus Search Engines

	Used Resources	Content Representation	Research Representation	Results Presentation
Directories	Accomplish them people	Manual classification	Implicit (browse through categories)	Created pages before consulting little exhaustive, very accurate.
Search Engines	Mainly in automated way through robots	Automated indexing	Explicit (keywords, operators, etc.)	Pages created Dynamically in each consult. Very exhaustive, little accurate.

Source: Martínez Méndez - 2000 - p.27.

Each one of the search engines uses secret algorithms[7] to order from the most to the least relevant the results returned to the users. The algorithm which generated more literature in the recent years was the one drawn and patented by the Google creators, in 1998, denominated 'PageRank'. It deals with a complex system based in web connections existing between the web pages. However the totality of the criteria that the system uses to calculate this data is reserved, it seems to be considerable the frequency of words appearance, its position in the text, the number of connections which lead themselves to a page or the importance of the page which receives and emit its vote. Like that a web page pointed twenty connections, apparently has less interest than the ones which are pointed one thousand connections. When making a research at the Google will place the first positions the pages which have a high PageRank and, that also, coincide with the thematic of the research.
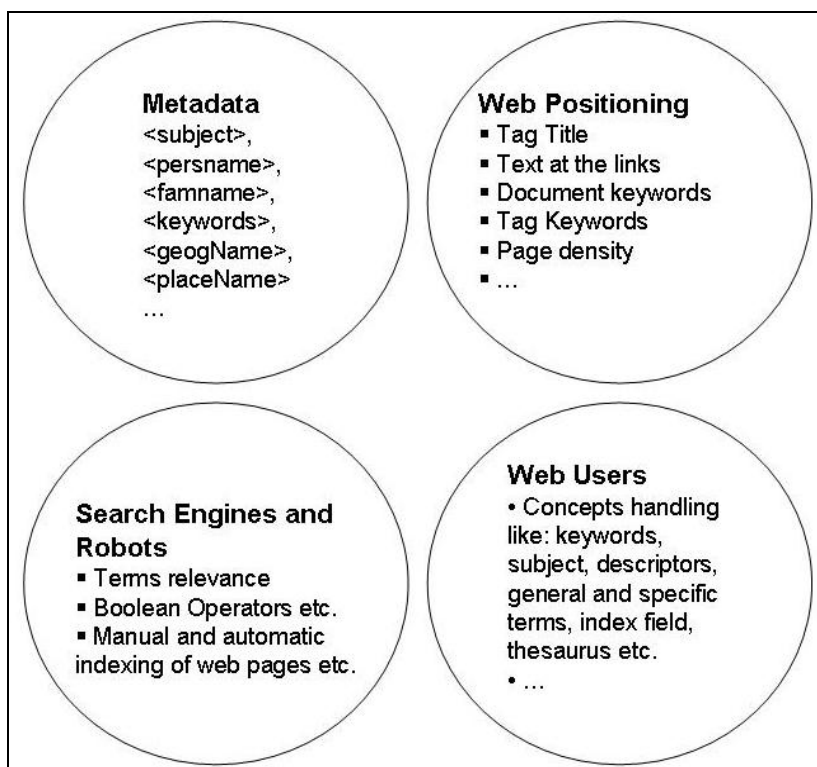
2.4 Users

The Internet use is converting each user into a *paradocumentalist* in potential. Unconsciously, the search users assimilate the terminology, the concepts and the practices which, until the middle of the 1990, were almost exclusive of information and documentation professionals.

The users who resort to the Internet assiduously to locate information (the last CD of his favorite singer, a book from the Municipal library, a house insurance, a monthly supermarket purchase etc) are acquainted with knowledge such as:

- To accomplish a research it is necessary to choose carefully the keywords. And these, the most specific the better.
- It is also possible to make use of phrases and not only of simple words.
- Some searchers allow you to put the text between quotation marks, in a way to obtain an exact agreement of compound-nouns or phrases.
- To restrict a research there are the boolean operators (and, or, not or no)
- It is possible to search information or a data in an specific area or tag (ex.: key-word: low cost flies).
- To specify a date or intervals reduce the number of documents recovered.

▪ The browse through subject list, a thematic classification or even a catalog to select the subject or the desidered term.

Finally, concepts such as *information recovering, relevant documents keywords, subjects and describers, specific term and general term, boolean operators, research in fields and tags, fields and tags index, keywords list, catalogs etc* turned popularized.

Metadata
<subject>,
<persname>,
<famname>,
<keywords>,
<geogName>,
<placeName>
...

Web Positioning
▪ Tag Title
▪ Text at the links
▪ Document keywords
▪ Tag Keywords
▪ Page density
▪ ...

Search Engines and Robots
▪ Terms relevance
▪ Boolean Operators etc.
▪ Manual and automatic indexing of web pages etc.

Web Users
• Concepts handling like: keywords, subject, descriptors, general and specific terms, index field, thesaurus etc.
• ...

Picture 3: Indexing Universe at the Internet.
Source: Prepared by the author.

CONCLUSIONS

The review of important elements which found the Internet web, for example, the metadata, the searchers, the users and the web positioning allowed us to find our starting hypothesis. It parts from the idea that these basic supports are impregnated, in higher or smaller measure, by the indexing, what leads to create an *Indexing Universe at the Web*, propitiated by the progressive extension of concepts and

practices propers of the indexers as consequence of the Internet popularization and extension.

REFERENCES

ALIMOHAMMADI, D. Meta-tag: a means to control the process of web indexing. **Online Information Review**, v.27, n.4, p.238-242, 2003.

CRAVEN, T. Variations in use of meta tag keywords by web pages in different languages. **Journal of Information Science**, v.30, n.3, p.268-279, 2004.

CRAVEN, T. Web authoring tools and meta tagging of page descriptions and keywords. **Online Information Review**, v.29, n.2, p.129-138, 2005.

GEORGE, D. The ABC of SEO: search engine optimization strategies. Morrisville, NC: Lulu Press, 2005.

GRAPPONE, J. Search engine optimization: an hour a day. San Francisco (CA): Sybex; Chichester: Wiley, 2006.

KENT, P. Search engine optimization for dummies. Indianapolis: Wiley, 2006.

MARCOS, M. C. et al. Evaluación del posicionamiento web en sistemas de información terminológicos online [en linea]. **Hipertext.net**, n.4, 2006. Available in: <http://www.hipertext.net>. Access in: 25 set. 2007.

MARTÍNEZ MÉNDEZ, F. J. **Propuesta y desarrollo de un modelo para la evaluación de la recuperación en Internet**. Murcia: Universidad de Murcia, 2000. (Thesis of doctored).

MÉNDEZ RODRÍGUEZ, E. **Metadatos y recuperación de información**: estándares, problemas y aplicabilidad en bibliotecas digitales. Gijón: Trea, 2002.

MERLO VEGA, J. A.; SORLI ROJO, A. El uso de metainformación en los webs de las bibliotecas españolas. In: FESABID. **Actas…** [S.l.p.]: Fesabid, 2000. p.155-164

SIROVICH, J.; DARIE, C. **Professional search engine optimization with PHP**: a developer's guide to SEO. Indianapolis: Wiley, 2007.

Prof. Dr. Isidoro Gil-Leiva
Facultad de Comunicación y Documentación
Universidad de Murcia
Murcia - Spain
isgil@um.es

Annex 1: Recommendations to a Good Web Positioning

Recommendations A	Recommendations B
1. Research the most used keywords to search web pages which offer the same that the own page. Use the tool <u>Wordtracker</u> or suggestion of keywords from the <u>Google Adwords</u>. 2. Chose the keywords to optimize the site pages; not only to the main page, but also to the internal pages. Try to make that each page covers from 2 to 4 keywords to the page content. 3. Include the keywords of each page in the titles, headings (H1, H2...) and the text. 4. Include the keywords in the text of the links of the internal browse, in a way that the keywords, to optimize a page concretely find themselves in all the links which point to the referred page. If you use images such as browse buttons, include the key words at the tag *ALT*. 5. Assure that all the web page is navigable by a navigator which has turned off '*javascript'* and *'flash'.* On the contrary, implement an alternative browse system with tags <noembed> or <noscript> or a map of the web page. 6. Obtain all the possible links of the web page with other web pages and, if possible, assure that in the text of links be included the most important keywords. This is the most important advice to obtain a good positioning. 7. Research with the keywords established in the web page in directories such as <u>DMOZ</u>, all the generalist directories which you find and all the directories specialized at the thematic of the web page. 8. Identify other web pages similar and ask for a link to the proper page. Search for links which point to competitive web pages, using the search chain "link:http://www.direccion.com". Thinking of new methods to obtain more links is very important. 9. Send the web page to all the possible <u>searchers</u>. 10. Review what <u>Google</u> considers *spam* not to be in risk of eliminating	1. **There are no shortcuts:** There is no easy and fast way to obtain good results. But on the contrary, it demands a lot of work. It is also necessary to have patience. The results do not come from night to day. 2. **Write the content well:** This is probably the most important thing which should be done, if you want to be found at the web. Grammatically correct, specific enough and always up-to-date. 3. **Think of writing correctly (good spelling):** When writing in English (in our case Portuguese), observe the differences between writing in American and British English (as the differences between the Portuguese from Brazil and from Portugal). 4. **Write the titles of the descriptive pages:** Make the titles of the pages simple, but descriptive and relevant, , in way that it becomes easier to the searchers know what each page deals about, and that the person when observing the search result, may determine if the web page contain what he is searching for. Some argue that it is one of the most important elements of the web page. Do not use the same title for all the documents. 5. **Use real headings:** Use the elements [HTML] h1 to h6 to the headings. 6. **Use friendly URLs to the searchers:** Avoid URLs generated automatically and which use the "query string" to make it possible to the server knows that the information brings the database. The search robots possibly have difficults with this kind of URLs. Use friendly URLs to the searchers and legible to the people. 7. **Make yourself linked:** There is no easier neither sustainable to solve this except to provide good content. The close links are very important to the SEO. 8. **Use a valid, semantic, fast and accessible marking [HTML]:** Validate the HTML and avoid the presentation marking [HTML].

the database.	Use a fast and clean marking [HTM] the most possible. Develop the content proportion in HTML [instead of graphics], this will turn the site faster and more attractive to the searchers. 9. **Aggregate the searchers to the site carefully:** Even excessively valorized, aggregating the site to the directories and searchers may be useful, especially if the site is new and, still, has not been caught by Google or other searchers. 10. **Do not mistake the searchers:** Do not use camouflage (cloaking, link farms), excess of keywords (keyword stuffing), alternative text with spam (alt text spamming) or other mistaken methods. They worked for a short period, but it is in risk of being afflicted, and what is worse, prohibited at the searchers. 11. **Avoid using marking (Frames):** They may cause problems to the person who accesses the site at the searchers. 12. **Take care with the detection of navigators:** if you need any type of navigator detection, assure that it functions when a robot of a searcher (or any other agent) arrives, if the robots can not enter, the site will not be found. 13. **Do not waste time with the meta tags**: Most of the searchers already do not give value the meta tags content, seen that they are exclusively used by the spammers. The keywords of <keywords> help very little, for this they are not worth.

Source: Recommendations A - <http://www.guiabuscadores.com/posicionamiento/posicionamiento-rapido.html>. Access in: Oct. 19, 2007.
Recommendations B - <http://uxespanol.blogspot.com/2006/03/principios-para-la-optimizacin-en.html>. Access in: Oct. 19, 2007.

NOTES

[1] Millán, José Antonio. El español y los buscadores. El País (Sección Opinión), Friday, September 22, 2006, p.15. In turn, the second text reproduced appeared in the section 'Cartas al Director' also from the news El País in March, 2007.

[2] Méndez Rodríguez, 2002, p.47.

[3] Méndez Rodríguez, 2002, p.60-61, who supports himself on Kashyap and Sheth, 2000, p.19-21.

[4] Merlo Verga and Sorli Rojo [2000] analyzed spanish libraries; Craven [2004 and 2005] studied the tags <keywords> of web pages of the 19 most present languages at the web, and another work, determined the effect of addition tools of web pages on the metatag keywords respectively; Alimohammadi (2004) calculated the presence of tags keywords in 346 web sites from Iran; and, last, Marcos et al. [2006] studied the positioning of ten web pages which hosted database and one of the criteria was the presence of <keywords>.

[5] Some publications printed recently are Sirovich and Darie [2007], kent [2006], Grappone [2006] or George [2005]. The web pages existing under this theme produce true vertex due to its quantity by what we offer only two examples: one in English and another in Spanish, which take as reference the first, but wide themselves in aspects.
http://www.optimi-seo-tion.com[Access in: 4.nov.2007]
http://www.guiabuscadores.com/posicionamiento/[Acess in: 4 Nov. 2007]

[6] Classifications taken out of:
http://www.web1marketing.com/blog/index.php/archives/top-10-seo-factors/ [Access in 19 out, 2007]
http://ezinearticles.com/?Top-15-On-Page-Good-SEO-Factor&id=294120 [Access in 19 out.2007]

[7] An algorithm is a set ordered and end of operations which make it possible to find the solution of a problem. The search engines use algorithms to locate, give value and subsequent, offer relevant web pages to the research accomplished.