

Module 10 Information & Research

Learning Unit 1 – Theory of Information Search and Retrieval

Authors

Alberto Hoces-García – Universidad de Murcia

Celia Chaín-Navarro – Universidad de Murcia

Licences

- This training material is subject to Creative Commons licenses. For training material, such as images, subject to another type of license, the license is explicitly stated
- This material is licensed under **Creative Commons Attribution-NonCommercial-ShareAlike 4.0 International License (CC BY-NC-SA 4.0)**.

To see a copy of this license, please visit

<https://creativecommons.org/licenses/by-nc-sa/4.0/deed.en>

Funding

- This training material has been developed within the framework of the **International Master for the Furniture Sector (IM-FUTURE)** project.
- The project is implemented within the framework of the **Cooperation for Innovation and the Exchange of Good Practices** Key Action and is funded by the **Erasmus+ Programme** of the **European Union**.

Learning outcomes

- This module will enable you to:
Understand the importance of materials and the connection between materials' properties and engineering technologies required to process them.
- Understand and critically evaluate aspects related to production organisation and optimisation
- Understand and critically evaluate aspects related to automation in furniture manufacturing
- Ability to understand and critically analyse the eco-sustainability aspects in furniture production

How to learn?

- This course could be taken using a “self directed” learning approach.
- For those attending “full time” courses a lecturer will be directing your learning using the learning materials available.
- For those enrolled on a “part time” basis a mixed approach could be employed.
- Please see below recommendations for each of the situations:

How to learn? Self directed approach

- We suggest the following sequence:
- Read the slides, watch the videos and reflect on the content.
- Read the additional text available where information is more detailed.
- If possible discuss with a colleague or register for a “Forum” and express your views.
- Test your understanding by taking the test at the end of the PowerPoint Presentation. Aim to score at least 70%
- If there are aspects that you find interesting and want to explore further access one of the texts recommended in the Bibliography

How to learn? Full time approach

- Attend all classes and the lecturer will provide explanations while showing the slides.
- Read the additional text available where information is more detailed.
- The lecturer will organise opportunities for you to discuss the content of the course so that you understand.
- Test your understanding by taking the test at the end of the PowerPoint Presentation. Aim to score at least 70%
- The lecturer will present a task for you to complete that will enable you to better understand the topic. The nature of the task depends on particular circumstances.
- If there are aspects that you find interesting and want to explore further access one of the texts recommended in the Bibliography

How to learn? Mixed approach

- Read the additional text available before attending the face to face classes.
- Attend all classes and the lecturer will provide explanations while showing the slides.
- The lecturer will organise opportunities for you to discuss the content of the course.
- Test your understanding by taking the test at the end of the PowerPoint Presentation. Aim to score at least 70%
- The lecturer will present a task for you to complete that will enable you to better understand the topic. The nature of the task depends on particular circumstances.
- If there are aspects that you find interesting and want to explore further access one of the texts recommended in the Bibliography

Why study Information & Research?

- To learn the basic concepts and principles related to the information search and retrieval, because you will be able to identify, select and acquire bigger and better information in less time.
- To learn the basic concepts and principles related to the scientific research methodology, because you will know which research method fits with your investigation, the stages and ethics that it must follow, where to search and what to write in a paper.
- Because you will be able to research in any field of your specialization.



Erasmus+



UNIT 1: Theory of the Information Search and Retrieval

MODULE 10:

SUBJECT 10.1: Information Search and Retrieval



Co-funded by the
Erasmus+ Programme
of the European Union

AFTER COMPLETION OF THIS MODULE, YOU WILL KNOW OR YOU WILL CAPABLE TO DO

Generic competences:

- *Adequate language skills:* Student is able to acquire, critically evaluate and creatively process information from the scientific literature databases, and other properly chosen sources.
- *Adaptability:* acknowledgement of the constant need to learn new skills and new concepts in a changing environment;
- *Logical reasoning abilities:* problem identification, creative search for solutions (both well-known ones and new ones), ability to follow logical inferences and elaborate formal reasoning in issues related with the information search and retrieval.



AFTER COMPLETION OF THIS MODULE, YOU WILL KNOW OR YOU WILL CAPABLE TO DO

Specific competences:

- Student is able to identify different model of information indexing;
- Student has knowledge of the methods of input and output information;
- Student has the ability to distinguish all categories of documents;
- Application of the tools of information search and retrieval;
- Student is aware of the importance of the normalization procedures;
- Student understand and can use impact indexes.



TOPICS 1. Theory of the Information Search and Retrieval

- 1. Introduction**
- 2. Information Search and Retrieval systems**
- 3. Documents search**
- 4. Questions**

1.1. Introduction

- Attention since the massive adoption of the Internet and information and communication technologies (ICT).
- Its main objective is that all the information retrieved is the relevant and necessary information to satisfy a search.



1.1. Introduction

- Factors that complicate the task of ISR:
 - The recovery system may not be easy to use or accessible.
 - The question may be incorrectly formulated.
 - The question may not correspond to the language used by the previous indexing of the information.
 - It is intended to recover all the information related to the question, not just one of the possible answers.
 - It is wanted to recover one of the possible answers, not all the information related to the question.
 - We do not want documents that do not respond to the question posed.
 - Recovered documents must be in the appropriate format.
 - The language of indexing and searching is different.



1.1. Introduction

- ISR requires knowledge of Information, Documentation, Linguistics, Computing, Statistics, etc.
- Several origins of the discipline are usually mentioned:
 - From the computational area, from Wiener's theory of cybernetics and Salton's subsequent contributions.
 - From the documentary area, as a result of the uniterms.
 - Some authors place the birth of ISR in ancient Greece, although they do not consider the current scientific paradigm.

1.1. Introduction

1.The stages according to Lesk

- Parallel with the ages of man of Shakespeare:
 - Childhood
 - Youthful
 - Adulthood
 - Maturity
 - Midlife crisis
 - Fullness
 - Retirement.



1.1. Introduction

- The childhood of the ISR is placed in 50's and 60's in which the construction of large-scale information systems, such as Dialog and BRS.
- The adult phase occurs in the 1970s, thanks to the arrival of text processors and other systems.
- Maturity occurs a decade later, with an immense increase in the available databases, the first OPACS and the first full-text online publications.



1.1. Introduction

- The crisis of middle age, approximately 45 years after the birth of ISR as a discipline, brings OPACs. The decade of the 90 'is marked by the birth of the Internet.
- The fullness of the discipline comes in the 2000s, thanks to the incorporation of multimedia information and the profusion of storage, reproduction, transmission, etc. devices.



1.1. Introduction



William Shakespeare

1.1. Introduction

2. The stages according to Salton

- Two stages: 1955-75 and from the 1975 onwards
 - First stage corresponded to the development of basic concepts and techniques
 - Second said techniques begin to be operational and new challenges arise due to the growth of electronic information.
- First phase
 - Generalization between the institutions of the use of computers.
 - Perception that there was a new horizon for the information.
 - In the decade of 1970 a great development takes place in this field, which also allowed the one of the ISR.



1.1. Introduction

3. The stages according to Moya

- Moya's theory is based on the development of the approaches and theoretical models on which the ISR discipline.
- The first stage locates its birth in 1940, at which time the problem of storage and retrieval of stored information was the object of the researchers.
 - Publication *As we may think* in 1945 produced an intense theoretical and practical work in the ISR, materialized in the 50s with the recovery of automated information through punched cards.
 - At this time, ISR is considered a static activity,
 - In the decade of the 1960s the first works on automation of documentary tasks arise.



1.1. Introduction

- The second stage begins its development during the decade of the 1970s, in which the first ISR systems developed specifically for certain institutions appear, which was the basis of the appearance of the first online public access catalogues (OPAC).



1.1. Introduction

- These systems are very appropriate for the retrieval of information through database management systems
- This operating principle is not so appropriate in the context of ISR, where users may have a more inaccurate idea of the information they are seeking, or as Belkin called it: "anomalous state of knowledge".

1.1. Introduction

- In the mid 70's, the development of information technology and telecommunications makes the ISR an interactive and dynamic process in which the user is perceived as an integral part of it.
 - Research during this decade focused on relevance feedback.
 - Theories and techniques of the previous decade were profiled and consolidated giving rise to the emergence of new models of ISR.
 - Information search services began to function.
 - It became evident that one of the problems of the ISR systems is the imprecise nature of the language.



1.1. Introduction

- The third stage would emerge as the 80's progressed for a series of differentiated reasons:
 - A notable increase in the number of documentary materials that needed to be indexed.
 - The cheapening of information storage devices led to the indexing of the full content of the documents.
 - The results of the investigations in information retrieval techniques, which promoted their integration into the operational ISR systems.
 - The generalization of the use of personal computers allowed users to begin to dispense with intermediaries when conducting searches.



1.1. Introduction

- Automation played a decisive role, since the new models and techniques were conceived for technology, but this did not imply in any way that traditional concepts linked to manual recovery lost their validity.
- In the late 1980s and early 1990s, many authors claimed more research based on real environments without excluding the user from the processes.



imfuture

1.2. Information Search and Retrieval systems

- There is no unanimous agreement among researchers when grouping the different models proposed to explain and represent it.

Four large groups:

- Boolean
 - Vectoral
 - Probabilistic
 - Cognitive.
- Another less widespread view is which reduces this division to two groups:
 - Traditional
 - Interactive.



1.2. Information Search and Retrieval systems

1. Boolean Model

- It is the most widely adopted and is used especially in commercial ISR systems.
- It is based on Boolean logic and set theory.
- Documents are searched by users through a series of questions conceived as a set of terms.
- This model is the simplest of all.

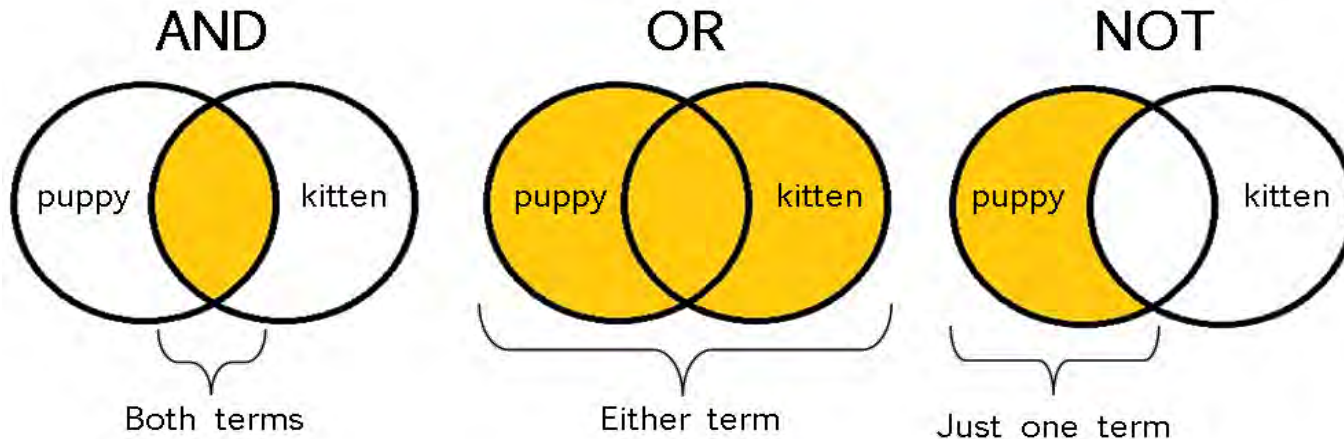


1.2. Information Search and Retrieval systems

- It presents a series of problems and deficiencies:
 - It is a system of data recovery rather than information
 - With the question/answer matching system used, it is usually recovered too many or too few documents.
 - Measures of precision are very low.

1.2. Information Search and Retrieval systems

- The operators of the system are common to the most basic computer languages:
 - AND, OR, NOT.
 - Parentheses to order by importance.



Boolean operators

1.2. Information Search and Retrieval systems

2. Vectoral Model

- Named because the representation of documents and questions is based on vectors.
- This model and parts from the theory of similarity, expressed by the degree of similarity between question and answer (document).
- The main difference (with Boolean) is that this range of documents is much more precise.



1.2. Information Search and Retrieval systems

- Its advantages are:
 - The scheme of the weight of the terms increases the recovery.
 - The strategy of approximation by similarity increases the possibility of retrieving documents that, although they do not coincide with the question, can be useful.
 - The cosine formula orders the documents according to the degree of similarity with the question.
- On the other hand, the main disadvantage is that it is assumed that the indexed terms are mutually independent, although dependency could also be a problem.



1.2. Information Search and Retrieval systems

3. SMART System

- Proposed by Salton, is distinguished from previous information retrieval systems in several features:
 - Use of fully automatic indexing methods.
 - It collects documents common to several classes.
 - It identifies the documents that can be recovered through the measure of similarity.
 - It includes automatic procedures to obtain information on the automatic procedures performed.



1.2. Information Search and Retrieval systems

- Each record or document is represented by "a vector of terms".
- The relevance feedback is based on the measure established by the users.
- In the same way that the similarity between question and document is established, it can also be done between two documents.

1.2. Information Search and Retrieval systems

- This reformulation is based on the following complementary operations:
 - The terms that occur in the documents previously identified as relevant by the users are added to the vectors of the original questions asked.
 - At the same time, terms that belong to documents classified as not relevant by users are cancelled or their importance is reduced.



1.2. Information Search and Retrieval systems

4. Probabilistic Model

- Is based on the fact that a probability of relevance (relative to a question) of a document is higher than irrelevance and exceeds a threshold, maximum or minimum value). The process works this way:
 - A question, documents and a cut value are given.
 - The probabilities that a document is relevant and irrelevant to a question can be calculated.



1.2. Information Search and Retrieval systems

- An optimal form of recovery is:
 - The documents are classified in order by their relevance.
 - The documents are recovered whose chances of relevance in the ranking exceed the cut-off value.
- The recovered documents are the first classified and the ordered list is then cut.
- There are several practical problems typical of this system:
 - It gives a correct value to the threshold for documents initially recovered.
 - The number of relevant documents retrieved per question is often too small for the probability to be considered reliable.



1.2. Information Search and Retrieval systems

5. Cognitive approach

- It groups several number of theories and methodological approaches; whose common point of view is the study of the user.
- It began at the end of the 1970s and it was influenced by cognitive science.
- Anomalous States of Knowledge (ASK)



1.2. Information Search and Retrieval systems

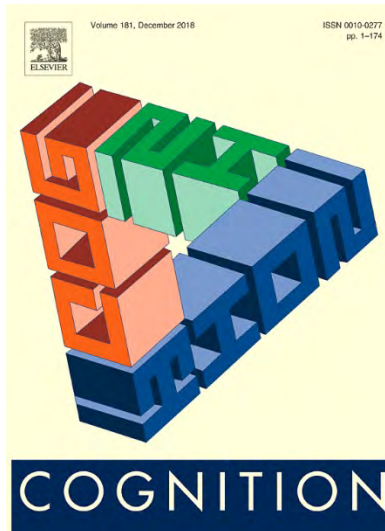
- "Tag effect", the tendency to label in verbal information through a few simple words or concepts. This generates two problems to the system:
 - The mechanisms of intermediation have problems to discern where to look for the information that the user needs.
 - The mechanism cannot find out whether this need is well or poorly expressed.
- The approach evolved later for several reasons:
 - A new concept of dynamic and complex relevance emerged.
 - This new perception produces a paradigm shift.
 - Other concepts changed in information retrieval: cognition and interaction.



1.2. Information Search and Retrieval systems

- Therefore, the two most important subprocesses of the cognitive approach are the processing of information and the need for information.
- On the other hand, an individual need for information is a function of the current state of knowledge of the individual.
- The information needs can be stable and well defined or variable, vague or poorly defined.

**Cognition - International Journal of
Cognitive Science**



imfuture

1.2. Information Search and Retrieval systems

6. Information recovery system

- The entry of documents into the system was preceded by a prior selection process through a series of internal and external sources
- Once they were in the documentary unit, they underwent a process of analysis.
- The theory is not easy to put into practice and documents are not always indexed or summarized.
- When it is possible to index and summarize the complete texts is when there is more chance that the recovery is more relevant.



1.2. Information Search and Retrieval systems

- Both in the full-text search and in the one based on descriptors or keywords, it is required that previously the indexing tasks have been adequately performed.
- This dependence conditions the way in which researchers must carry out our searches.
- Even by the automation of thesaurus indexing errors can occur and the intervention of a person who reviews the results is necessary.
- The review of the literature on the recovery of information makes it clear that, depending on the origin of the author, it is emphasized or focused more on a part of the process.

1.2. Information Search and Retrieval systems



"Analogic" information

The logo for 'imfuture' features a stylized 'i' composed of three horizontal bars in blue, green, and red. This is followed by a large, bold, dark brown 'M' and the word 'future' in a lowercase, dark brown, sans-serif font.

1.3. Documents search

- A document is an article, a book chapter, a book, a painting, a letter, a video, an image, etc.
- Is the question of the user a document in itself?



1.3. Documents search

- A question can be considered, or not, as a document that an information system can handle.
- If the objective is to define a match between a given question and those documents that the user would like to retrieve to answer that question, the matching process is complicated by the fact that the question and the documents can have different forms.

1.3. Documents search

- We expect an exact answer to the question: to find the document or documents that we are looking for easily and in the first place.
 - Only possible when the question that is formulated is identified by the database as answerable.
 - Most of the texts and images of the databases cannot be adapted to this system.
 - A system is capable of processing full-text documents, but it may be that the question that is asked has mixed characteristics.



1.3. Documents search

- Two ways of a question: a formulation in natural language and as a set of meaningful terms in the content.
- A standard model for developing questions from a series of terms is the Boolean model.



Digital searching

1.3. Documents search

- In Boolean questions, the use of the AND connector requires that both terms be present in the retrieved documents,
- OR requests that at least one of them be.
- Inclusive OR and an exclusive NOR.
- Despite the apparent simplicity, this type of Boolean questions presents several problems >

1.3. Documents search

- In Boolean questions there is no way to determine the weight of the terms.
- Irrelevant information is retrieved because the question is wrong.
- The order of preference of the logical connectors.

A OR B AND C



A OR (B AND C)



(A OR B) AND C

1.3. Documents search

- Most researchers do not know the rules of Boolean algebra and the particular priorities established by each information retrieval system.
- Each Boolean question can be included within normal disjunctive form or normal conjunctive form, each of which provides a standard to facilitate the process:
 - The normal disjunctive form has three levels of expression:
 - a) Terms.
 - b) Conjunctions.
 - c) Disjunctions.
 - A question in normal disjunctive form can have one or more disjunctions.

(small AND simple AND NOT fast) OR (manageable AND nice) OR
(powerful AND NOT heavy)



1.3. Documents search

- The questions in normal conjunctive form are defined in a similar way, but the roles of AND and OR exchanged. A question in this form would be:

(small OR simple NOR fast) AND (manageable nice OR) AND (powerful
NOR heavy)

- The process of transforming a Boolean question into a normal disjunctive form or into a conjunctive is called normalization.
- The fifth problem associated with Boolean logic is the control of the size and quantity of documents retrieved.



1.3. Documents search

- Questions in the vector model work differently. In it, each document is represented by a vector or ordered list of terms.
 - The absence or presence of the term can be represented by 0 and 1.
 - The use of the weight of the vector or value assigned to each term can be used to evaluate the similarity.
 - Since not all terms can be represented in all documents, the contents and their corresponding weights are represented.
 - The key to this model is maintaining dimensional compatibility.



1.3. Documents search

1. Documents output

- The output of the information retrieval system is rarely the exact set of documents the user wants to answer its question.
- The output is limited by the data available in the bases.
- The output of the system can offer the following information on the question posed:
 - Identification code of a document.
 - One or several bibliographical references.
 - Secondary information of the type of bibliographic reference with summary and / or keywords.
 - The complete document.



1.4. Questions

- 1) What is the difference between the stages of Lesk and Salton with those of Moya? Discuss the most important concepts developed during the 20th century for today's ISR.
- 2) Advantages and disadvantages of the Boolean and Vectoral models. Look for a search engine and explain in which model is based.
- 3) Make 3 different searches at Google Scholar using the advanced search and compare the results. Are the fields related to the Boolean model connectors?

Key references

- Belkin, N. J., Oddy, R. N., & Brooks, H. M. (1982). ASK for information retrieval: Part I. Background and theory. *Journal of documentation*, 38(2), 61-71.
- Buckley, C., Salton, G., & Allan, J. (1993, March). Automatic retrieval with locality information using SMART. In *Proceedings of the First Text REtrieval Conference TREC-1* (pp. 59-72).
- Chaín Navarro, C. (2004). *Técnicas y métodos de recuperación de la información*. Diego Marín
- De Mey, M. (2012). *The cognitive paradigm: Cognitive science, a newly explored approach to the study of cognition applied in an analysis of science and scientific knowledge* (Vol. 1). Springer Science & Business Media.
- Lesk, M. (1995). *The seven ages of information retrieval*. Universal dataflow and telecommunications core programme.
- Salton, G., & Harman, D. (2003). *Information retrieval*. John Wiley and Sons Ltd.

Module 10 Information & Research

Learning Unit 2 - Documents

Authors

Alberto Hoces-García – Universidad de Murcia

Celia Chaín-Navarro – Universidad de Murcia

Licences

- This training material is subject to Creative Commons licenses. For training material, such as images, subject to another type of license, the license is explicitly stated
- This material is licensed under
Creative Commons Attribution-NonCommercial-ShareAlike 4.0 International License
(CC BY-NC-SA 4.0).

To see a copy of this license, please visit

<https://creativecommons.org/licenses/by-nc-sa/4.0/deed.en>

Funding

- This training material has been developed within the framework of the **International Master for the Furniture Sector (IM-FUTURE)** project.
- The project is implemented within the framework of the **Cooperation for Innovation and the Exchange of Good Practices** Key Action and is funded by the **Erasmus+ Programme** of the **European Union**.

Learning outcomes

- This module will enable you to:
Understand the importance of materials and the connection between materials' properties and engineering technologies required to process them.
- Understand and critically evaluate aspects related to production organisation and optimisation
- Understand and critically evaluate aspects related to automation in furniture manufacturing
- Ability to understand and critically analyse the eco-sustainability aspects in furniture production

How to learn?

- This course could be taken using a “self directed” learning approach.
- For those attending “full time” courses a lecturer will be directing your learning using the learning materials available.
- For those enrolled on a “part time” basis a mixed approach could be employed.
- Please see below recommendations for each of the situations:

How to learn? Self directed approach

- We suggest the following sequence:
- Read the slides, watch the videos and reflect on the content.
- Read the additional text available where information is more detailed.
- If possible discuss with a colleague or register for a “Forum” and express your views.
- Test your understanding by taking the test at the end of the PowerPoint Presentation. Aim to score at least 70%
- If there are aspects that you find interesting and want to explore further access one of the texts recommended in the Bibliography

How to learn? Full time approach

- Attend all classes and the lecturer will provide explanations while showing the slides.
- Read the additional text available where information is more detailed.
- The lecturer will organise opportunities for you to discuss the content of the course so that you understand.
- Test your understanding by taking the test at the end of the PowerPoint Presentation. Aim to score at least 70%
- The lecturer will present a task for you to complete that will enable you to better understand the topic. The nature of the task depends on particular circumstances.
- If there are aspects that you find interesting and want to explore further access one of the texts recommended in the Bibliography

How to learn? Mixed approach

- Read the additional text available before attending the face to face classes.
- Attend all classes and the lecturer will provide explanations while showing the slides.
- The lecturer will organise opportunities for you to discuss the content of the course.
- Test your understanding by taking the test at the end of the PowerPoint Presentation. Aim to score at least 70%
- The lecturer will present a task for you to complete that will enable you to better understand the topic. The nature of the task depends on particular circumstances.
- If there are aspects that you find interesting and want to explore further access one of the texts recommended in the Bibliography

Why study Information & Research?

- To learn the basic concepts and principles related to the information search and retrieval, because you will be able to identify, select and acquire bigger and better information in less time.
- To learn the basic concepts and principles related to the scientific research methodology, because you will know which research method fits with your investigation, the stages and ethics that it must follow, where to search and what to write in a paper.
- Because you will be able to research in any field of your specialization.



Erasmus+



UNIT 2: Documents

MODULE 10:

SUBJECT 10.1: Information Search and Retrieval



Co-funded by the
Erasmus+ Programme
of the European Union

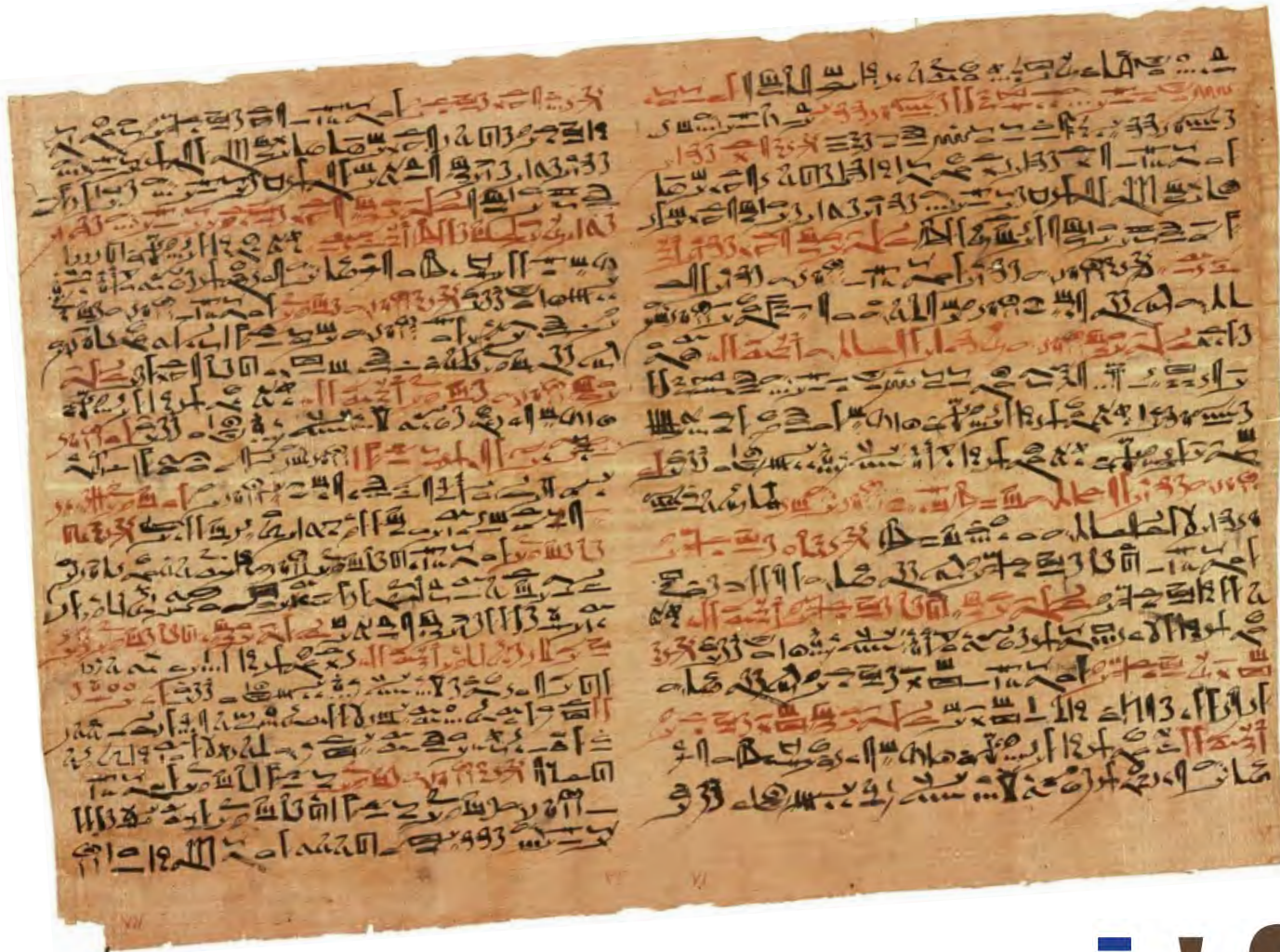
TOPICS 2. Documents

1. The primary documents
2. The secondary documents
3. Questions

2.1. The primary documents

- The primary documents are those in which the original information is exposed.
- Science advances through the publication of these documents.
- The first type of document that fulfils the four characteristics given to the primary documents is the periodical publication:
 - Specialized
 - Current
 - Agile
 - Synthetic information.

2.1. The primary documents



Primary document in papyrus

2.1. The primary documents

1. Periodical literature

- Scientific papers are the way in which most investigations are published.
- Quicker than the monographs, greater specialization, less extension and a more technical language.
- Papers are published in specialized journals.
- In addition to articles, scientific journals usually contain other types of information:
 - Revisions.
 - Short news.
 - Reviews.

2.1. The primary documents

- These three types of information cannot be considered as primary, since they do not have original information, but it is synthesized or revised.
- To consider the originality, quality and validity of the papers they are subject to evaluation by a committee.
- Journals have mandatory publication rules for the presentation. The information that each article must contain:
 - Title
 - Author / s
 - Institution and email
 - Summary
 - Keywords
 - Title, summary and keywords in other languages, usually English if the main journal or the article are in another language.

2.1. The primary documents

- The content itself is initiated by an introduction.
- The methodology used by the author or the team is reproduced below.
- The extension of the articles varies considerably based on the investigation carried out and its importance.
- Due to peer review processes, the process of sending a paper to a journal may delay its publication more than six months.

The two most known scientific journals

 **iMfuture**



2.1. The primary documents

2. Thesis

- The necessary works to obtain certain certifications such as degree, master or doctorate offer some of the first investigations of the new authors.
- The thesis agglutinate almost all the available content on the subject (especially those of doctor degree) and usually are the main source of information a concrete subject.
- Very specialized works aimed at a very specific public.

2.1. The primary documents

3. Minutes of communications presented at conferences

- Research advances or short articles on research in progress or recently completed are published in the minutes.
- They have recently been losing the most interesting characteristic, the rapid publication, in search of peer review processes that ensure their quality.

2.1. The primary documents

4. Reports

- Much of the applied research is written in the form of technical or research reports.
- Origin in commercial, military or projects of work teams.
- Restricted content because of commercial or military interest cause an almost non-existent diffusion.

2.1. The primary documents

5. Patents

- The patents, granted by the World Intellectual Property Organization (WIPO), are the official records that certify the intellectual property of an object, advance, discovery, technique, etc.

2.1. The primary documents

6. Books and monographs

- The book format is the best known for obtaining information due to its great tradition.
- Its editorial characteristics and its extension make it less updated than the periodical publications
- They offer complete information about a specific topic in a simpler way; which facilitates its diffusion.

2.1. The primary documents

7. Commercial literature

- The commercial literature contains information not accessible from other types of publications, such as prices or designs.

8. Standards

- They are the result of the normalization of research and work to create a uniform product; offering primary information despite being based on other documents for its preparation.

2.1. The primary documents



Conferences are usually published in minutes

2.2. The secondary documents

- Secondary documents are those that collect data and information from primary documents, such as reference works.

1. The development of instruments

- There are different types of secondary documents:
 - That start from the analysis of the primary ones and that always consign to other documents to find the information.
 - That offer information directly, since they define terms and/or present ideas.
 - Directories, which reflect particularities of the two previous ones.

2.2. The secondary documents

- They are also called reference works since they offer this type of information (references).
- Critical revisions or reviews may be included in this division.
- Repertoires of reference works are called tertiary documents.

2.2. The secondary documents

2. Secondary documents generated from the analysis of the primary

- They are those that always refer to other documents to find the required information.
- The difference between them lies in the depth of information collected, easier in the first three and more laborious in the last two:
 - Bibliographic repertoires.
 - Catalogues of libraries, collectives and bookstores.
 - Bulletins, or repertoires, of summaries.
 - Bulletins, or repertoires, of indexes.
 - Bulletins, or repertoires, of abstracts or "abstracts".

2.2. The secondary documents

3. Bibliographic repertoires or bibliographies

- These are lists of bibliographic references of primary documents on more than one scientific field that are used to keep abreast of what has been published.
- Since the term bibliography includes several senses:
 - Science of books,
 - Complete or selective list of documents on a topic or
 - periodic list of recently published documents;

2.2. The secondary documents

- It is a different reality with different types of bibliographies:
 - a) International or universal.
 - b) National general bibliographies.
 - c) Specialized bibliographies.
 - d) Bibliographies of serial publications, catalogues or serials repertoires.
- The bibliographies are differentiated according to:
 - Their periodicity.
 - The extension of the documents registered.
 - The exposition order of the content.

2.2. The secondary documents

4. Catalogues

- A catalogue is a list of all information documents in an institution with a specific order of presentation: by author, subject, etc.
- In the libraries there is usually an inventory of funds together with a catalogue or OPAC of public consultation.
- The collective catalogues are a list of the catalogues of several libraries.
- Commercial catalogues compile the works of bookstores and publishing houses for sale.

2.2. The secondary documents

5. Repertoires of summaries

- Also called bulletins of summaries.
- Secondary documents that collect summaries of journals, usually specialized in some discipline.



2.2. The secondary documents

6. Permuted indexes

- Most them are done automatically.
- They facilitate the consultation of bibliographic bulletins or summaries to which they usually accompany.
- Although some of them may include summaries of the indexed journals, they are considered within the repertoires of indexes if:
 - The main part of their contents are indexes, more elaborate than those of bulletins of summaries.
 - Its intention is to facilitate a retrospective search.
 - Its frequency is usually annual.

2.2. The secondary documents

words on left

Permuted Index + words on right

Line

Word

The quick

brown fox

0

2

jumped over the

fence

1

3

The quick brown

fox

0

3

jumped over the fence

1

0

jumped

over the fence

1

1

The

quick brown fox

0

1

jumped over

the fence

1

2

The quick brown fox

0

0

KWIC permuted index

- There are several kinds of permuted indexes:

- KWIC and KWIT.
- KWOC.
- Cross-indexes
- Cumulative indexes.
- Citation rates.
- Impact indexes.

2.2. The secondary documents

7. Repertories of abstracts

- They allow the researcher to know in a synthetic way the content of a document without having to read it completely.

8. Dictionaries and lexicographical works

- A dictionary is a collection of words in a language, classified in alphabetical order and explained in the same language.
- It is also a collection of words from one language translated into another.
- Those dedicated to complex terms, technical or dedicated to a profession or sector are called glossaries.

2.2. The secondary documents

- Lexicons or vocabularies are a type of dictionaries that cover a specific field.
- The nomenclatures differ from them in their normalization.
- Another type of standardization is that elaborated in the thesauri or standardized lexicons.
- Different because they do not define terms are the biographical dictionaries.



2.2. The secondary documents

9. Encyclopaedias

- The encyclopaedias are responsible for presenting the state of knowledge at the time of publication.
- They do not cover all the concepts on a subject but a selection of essentials, and in a more extensive way than the dictionary.
- Publication of paper encyclopaedias is becoming less common.
- The Internet has made possible the appearance of encyclopaedias written collaboratively or wikis.



2.2. The secondary documents

10. Manuals and yearbooks

- Primary documents that give an overview of the essential contents of a scientific or thematic discipline.
- Some of them are published on an annual basis.
- There are different types of yearbooks:
 - a)Chronologies.
 - b)Annual balances.

11. Directories

- In the directories there are indications about documents, organisms or people of a specific period or field.



2.2. The secondary documents

12. Institution directories

- These are specialized directories in which a brief description of the organizations that work in a certain field is offered.
- Yearbooks can also report on administrative services.

13. People directories

- People directories provide information about members of professional organizations and active associations.

14. Research project directories

- They include research projects, both ongoing and completed, organized by specialties, by the agency they depend on or that finances them.



2.2. The secondary documents

15. Thesis catalogues

- They collect information about the thesis in progress, presented or published by universities and other academic institutions.

16. Contract guides

- They give access to information about research funded by an agency or by all the agencies in a country that sponsors the activities of the research groups.

2.3. Questions

- 1) Discuss the relevance of the differences between a primary and a secondary document for the research in the furniture areas of business, design and management.
- 2) Write an abstract of 150 words of an imaginary paper related with your research area.
- 3) Explain the order of the kind of documents you would search in if you had to write a thesis.

Key references

Chaín Navarro, C. (2004). *Técnicas y métodos de recuperación de la información*. Diego Marín.

_____. (2004). Coincidencia y equiparación en los modelos de recuperación de información. *Documentación de las Ciencias de la Información*, 27, 219-236.

_____. (1995). *Introducción a la gestión y análisis de recursos de información en ciencia y tecnología*. EDITUM.

Cordón García, J. A., Gómez Díaz, R., Alonso Arévalo, J., & López Lucas, J. (2010). *Las nuevas fuentes de información: información y búsqueda documental en el contexto de la web 2.0*. Pirámide.

Farace, D., & Schöpfel, J. (2017). *Grey literature in library and information studies*. KG Saur.

Module 10 Information & Research

Learning Unit 3 - Normalization

Authors

Alberto Hoces-García – Universidad de Murcia

Celia Chaín-Navarro – Universidad de Murcia

Licences

- This training material is subject to Creative Commons licenses. For training material, such as images, subject to another type of license, the license is explicitly stated
- This material is licensed under **Creative Commons Attribution-NonCommercial-ShareAlike 4.0 International License (CC BY-NC-SA 4.0)**.

To see a copy of this license, please visit

<https://creativecommons.org/licenses/by-nc-sa/4.0/deed.en>

Funding

- This training material has been developed within the framework of the **International Master for the Furniture Sector (IM-FUTURE)** project.
- The project is implemented within the framework of the **Cooperation for Innovation and the Exchange of Good Practices** Key Action and is funded by the **Erasmus+ Programme** of the **European Union**.

Learning outcomes

- This module will enable you to:
Understand the importance of materials and the connection between materials' properties and engineering technologies required to process them.
- Understand and critically evaluate aspects related to production organisation and optimisation
- Understand and critically evaluate aspects related to automation in furniture manufacturing
- Ability to understand and critically analyse the eco-sustainability aspects in furniture production

How to learn?

- This course could be taken using a “self directed” learning approach.
- For those attending “full time” courses a lecturer will be directing your learning using the learning materials available.
- For those enrolled on a “part time” basis a mixed approach could be employed.
- Please see below recommendations for each of the situations:

How to learn? Self directed approach

- We suggest the following sequence:
- Read the slides, watch the videos and reflect on the content.
- Read the additional text available where information is more detailed.
- If possible discuss with a colleague or register for a “Forum” and express your views.
- Test your understanding by taking the test at the end of the PowerPoint Presentation. Aim to score at least 70%
- If there are aspects that you find interesting and want to explore further access one of the texts recommended in the Bibliography

How to learn? Full time approach

- Attend all classes and the lecturer will provide explanations while showing the slides.
- Read the additional text available where information is more detailed.
- The lecturer will organise opportunities for you to discuss the content of the course so that you understand.
- Test your understanding by taking the test at the end of the PowerPoint Presentation. Aim to score at least 70%
- The lecturer will present a task for you to complete that will enable you to better understand the topic. The nature of the task depends on particular circumstances.
- If there are aspects that you find interesting and want to explore further access one of the texts recommended in the Bibliography

How to learn? Mixed approach

- Read the additional text available before attending the face to face classes.
- Attend all classes and the lecturer will provide explanations while showing the slides.
- The lecturer will organise opportunities for you to discuss the content of the course.
- Test your understanding by taking the test at the end of the PowerPoint Presentation. Aim to score at least 70%
- The lecturer will present a task for you to complete that will enable you to better understand the topic. The nature of the task depends on particular circumstances.
- If there are aspects that you find interesting and want to explore further access one of the texts recommended in the Bibliography

Why study Information & Research?

- To learn the basic concepts and principles related to the information search and retrieval, because you will be able to identify, select and acquire bigger and better information in less time.
- To learn the basic concepts and principles related to the scientific research methodology, because you will know which research method fits with your investigation, the stages and ethics that it must follow, where to search and what to write in a paper.
- Because you will be able to research in any field of your specialization.



Erasmus+



UNIT 3: Normalization

MODULE 10:

SUBJECT 10.1: Information Search and Retrieval



Co-funded by the
Erasmus+ Programme
of the European Union

TOPICS 3. Normalization

1. Types of standards
2. Standardization bodies
3. Previous attempts to control information in Science and Technology
4. The rules for bibliographic control
5. The standardised numbers of identification of publications
6. Bibliographic references
7. Questions

3. Normalization

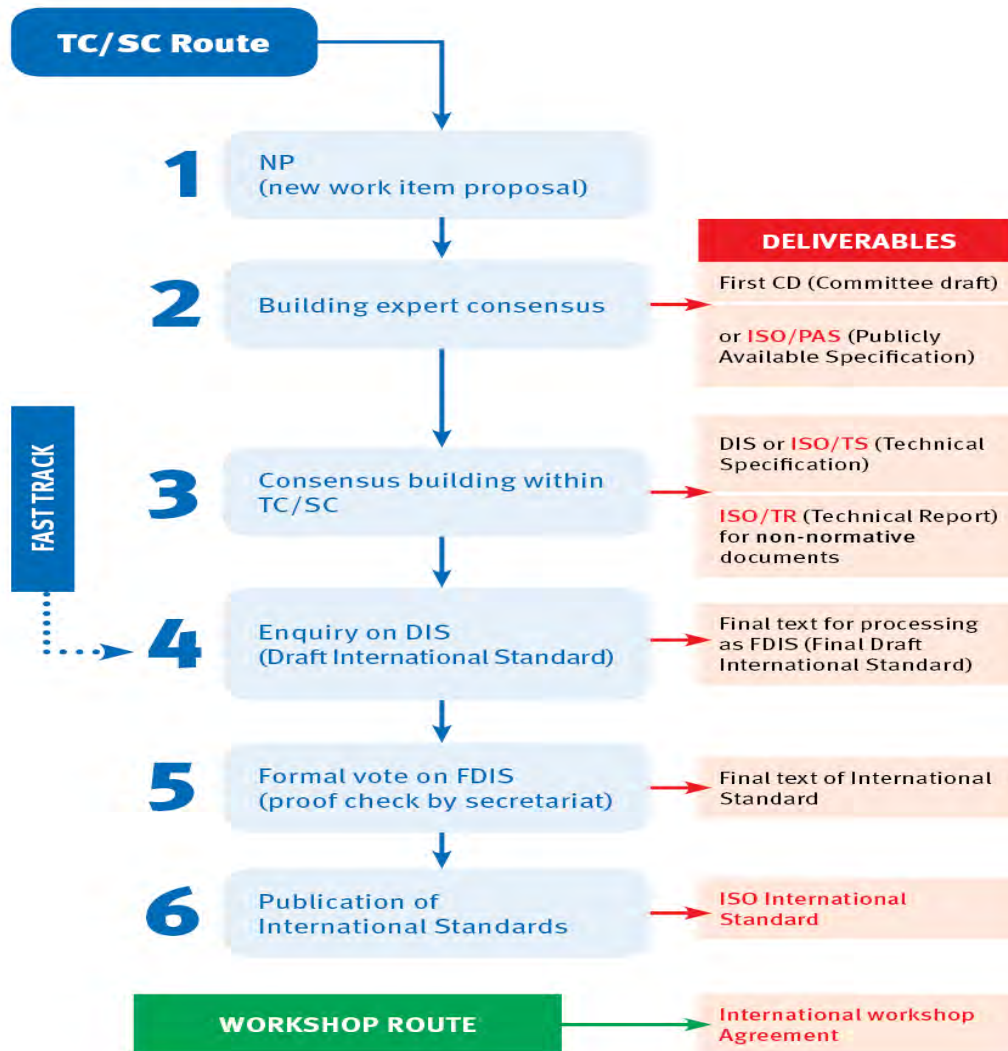
- Standardization is the collective activity through which standards are established.
- A standard is a guideline that has a rule value, generally indicative and sometimes mandatory.
- Its purpose is to define the characteristics that an object must have and those that its use must have, as well as those of a procedure and / or a method.
- There are norms to typify material objects (formats), and for intellectual operations (documentary analysis)

3. Normalization

- Standardization is very important for the creation and organization of international information systems within the academic and scientific field.
- Normalization affects the information retrieval because it provides the researcher a set of rules for the development of his/her work.
- UNESCO planned to launch a program to channel all these aspirations at the international level: UNISIST within the GIP (General Information Program).
- The work of most of the international organizations is focused today on:
 - The search of common formats of communication in all areas.
 - The establishment of hierarchies and lists that allow to integrate and structure knowledge.
 - The elaboration of international standards for the description of the materials of the documentary institutions.



3. Normalization



**ISO Standards
development process**

3.1. Types of standards

- Dimension rules
- Quality standards
- Definitions, lexical rules, standardized terms and symbols.
- Standard procedures and procedures.
- The fields of application of the standards within the ISR scope are:
 - The presentation of documents.
 - Bibliographic control.
 - The bibliographic reference.
 - The transliteration.
 - The terminology.
 - The description of the content.
 - The realization of catalogues and files.
 - The premises and the equipment.
 - Reproduction.
 - Computing.
 - The transmission of data in networks.
 - The management.
 - Statistics.
 - Information systems.



3.2. Standardization bodies

- The standards are developed within standardization bodies by specialists in each field. The organisms can be:

1. National:

- a. Standardization offices and services.
- b. A national standardization body.
- c. A higher authority related to the administration.

2. International:

- There are several standardization bodies.
- The ISO International Organization for Standardization represents the global standardization body.



3.3. Previous attempts to control information in Science and Technology

1. International policies

- Universal Availability of Publications (UAP) is an objective and a program.
- The development of the UAP program has been supported by UNESCO, the Deutsche Forschungsgemeinschaft, the British Library and other organizations and institutions.

3.3. Previous attempts to control information in Science and Technology

- Making it effective requires:
 - The information needs of the users to be analysed.
 - That editing and distribution programs of the editors in each of the countries are adjusted to the demands of national and foreign users.
 - The book trade channels to be established or improved to ensure users the rapid availability of publications.
 - Libraries and related entities to develop effective policies and procedures for the acquisition of material.
 - Libraries to develop procedures and programs for the supply of original publications or copies to the remote user as well as to the local one.
 - Each country to ensure the permanent conservation of at least one copy of the publications acquired by its libraries and archives.
 - Each country to accept full responsibility for the acquisition, supply through the loan and conservation of all its publications to meet the present and future needs of its own inhabitants and other countries.
- The professional potential in the field of libraries, documentation and information to be developed and improved.
- The reading technique and the easy handling of information to develop in individuals from the earliest stages.
- Users are committed to the improvement plans.
- New information technologies should be used appropriately.



3.3. Previous attempts to control information in Science and Technology

- In conclusion to the International Congress on Universal Availability of Publications, recommendations were issued:
 - a) General recommendations, addressed jointly to international organizations, governments and information professionals;
 - b) Recommendations addressed to the Member States of UNESCO;
 - c) Recommendations addressed to UNESCO and the FIAB, and >

3.3. Previous attempts to control information in Science and Technology

d) Recommendations addressed to the World Book Congress.

- There were a total of 48 recommendations or proposed lines of action.
- Other core programs for IFLA related to the UAP are:
 - Universal Bibliographic Control (UBC).
 - Marc International Program (IMP).
 - Preservation and Conservation (PAC).
 - International Data Flow (TDF).
 - Progress of Librarianship in the Third World (PLTW).

3.3. Previous attempts to control information in Science and Technology



It is hard to control the immense amount of information without normalization

3.4. The rules for bibliographic control

1. ISBD

- FIAB organized in 1969 in Copenhagen an international meeting of cataloguing experts, and from it the ISBDs emerged.
- The working group began to meet that same year with the purpose of offering universal use standards:
- In 1971 the first standards appeared, which were soon adopted by three national bibliographies (those of Great Britain, France and the then Federal Republic of Germany).
- The success of ISBDs (M) led to the development of other standards for different types of materials.
- The idea of creating norms of general application for all types of materials was forged. The ISBDs (G) appeared, agreeing that the specialized ISBDs had to adapt to the guidelines established in it

3.4. The rules for bibliographic control

2. ISDS

- The ISDS, or International Data System on Serial Publications, is an intergovernmental organization created within the UNISIST program.
- Its purpose is to create the ISSN, a reliable database of all serial publications in the world.
- The ISDS format is based on the serial publications established in the MARC format

3.5. The standardized numbers of identification of publications

1. ISBN

- The ISBN is a standard international number for the identification of books and monographs.
- Its composition and structure are governed by the ISO 2108-1978 NORM.
- Its purpose is to coordinate and standardize the use of the number of books.
- An ISBN consists of ten digits composed of four parts, separated by a hyphen or by blanks:
 - An identifier group.
 - Editor's ID.
 - Title identifier.
 - Control digit.

3.5. The standardized numbers of identification of publications

ISBN 978-3-16-148410-0

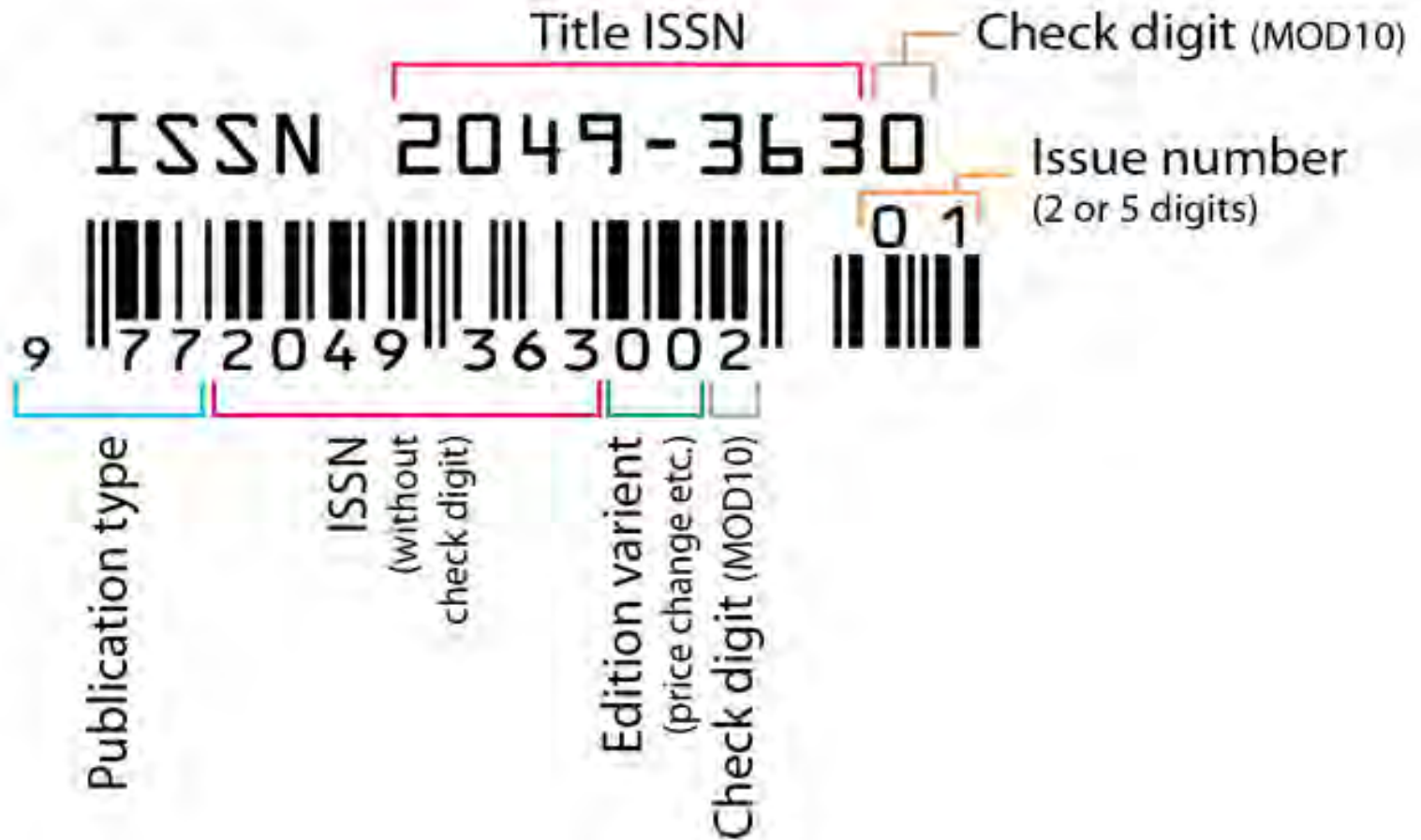


3.5. The standardized numbers of identification of publications

2. ISSN

- The ISSN (International Standard Serial Number) is a numeric code for the identification of serial publications.
- The ISSN is composed by eight figures and does not incorporate any other meaning than the identification of the serial.

3.5. The standardized numbers of identification of publications



3.5. The standardized numbers of identification of publications

3. BIBLID

- The Biblid is a standardized code.
- Its main function is to provide, on the first page of each article, a univocal identifier normalized in a directly readable coded form.
- The Biblid is regulated in the ISO 9115-1987 standard.

3.5. The standardized numbers of identification of publications

- The Biblid consists of a code identifier and the code itself. There are two types:
 - The Biblid for serial publications.
 - The Biblid for books.
- The Biblid is designed to be used by manual and automated means.
- The Biblid format requires an uninterrupted sequence of characters, with no blank spaces.

3.5. The standardized numbers of identification of publications

1. Biblid for serial publications

- a) The first element of the Biblid for books is the ISSN number.
- b) The second part is the year of publication.
- c) The next element is the fascicle indication.
- d) The fourth component is paging.

3.5. The standardized numbers of identification of publications

2. Biblid for books

- a) The first element of the Biblid for books is the ISBN number.
- b) The second element is the year of publication.
- c) The third element is the pagination.

3.6. Bibliographic references

- One of the most important aspects of standardization in its application to research is the control of the references made in the works.
- There are international standards, which unfortunately are not of global use.
- The rules of more extended use are:
 - ISO.
 - Vancouver.
 - Chicago.
 - APA.
 - MLA.
- There are also different ways to present them in papers:
 - Quotes at the footer.
 - Harvard style quotes, or author-date system.
 - Mixed between both.

2.3. Questions

- 1) Look for a standard in the furniture industry and discuss how it can affect to the research in this field.
- 2) Identify the international bodies of standardization in the furniture industry
- 3) Explain the parts of the ISBN of a book about furniture design.

Key references

Chaín Navarro, C. (2004). *Técnicas y métodos de recuperación de la información*. Diego Marín.

_____. (2000) *Técnicas documentales aplicadas a la investigación*. Murcia: ICE.

_____. (1995). *Introducción a la gestión y análisis de recursos de información en ciencia y tecnología*. EDITUM.

Gorman, M. (2004). Authority control in the context of bibliographic control in the electronic environment. *Cataloging & Classification Quarterly*, 38(3-4), 11-22.

Smith, L. C., & Wong, M. A. (Eds.). (2016). *Reference and Information Services: An Introduction*. ABC-CLIO.

Module 10 Information & Research

Learning Unit 4 – Information Retrieval from the Web

Authors

Alberto Hoces-García – Universidad de Murcia

Celia Chaín-Navarro – Universidad de Murcia

Licences

- This training material is subject to Creative Commons licenses. For training material, such as images, subject to another type of license, the license is explicitly stated
- This material is licensed under
Creative Commons Attribution-NonCommercial-ShareAlike 4.0 International License
(CC BY-NC-SA 4.0).

To see a copy of this license, please visit

<https://creativecommons.org/licenses/by-nc-sa/4.0/deed.en>

Funding

- This training material has been developed within the framework of the **International Master for the Furniture Sector (IM-FUTURE)** project.
- The project is implemented within the framework of the **Cooperation for Innovation and the Exchange of Good Practices** Key Action and is funded by the **Erasmus+ Programme** of the **European Union**.

Learning outcomes

- This module will enable you to:
Understand the importance of materials and the connection between materials' properties and engineering technologies required to process them.
- Understand and critically evaluate aspects related to production organisation and optimisation
- Understand and critically evaluate aspects related to automation in furniture manufacturing
- Ability to understand and critically analyse the eco-sustainability aspects in furniture production

How to learn?

- This course could be taken using a “self directed” learning approach.
- For those attending “full time” courses a lecturer will be directing your learning using the learning materials available.
- For those enrolled on a “part time” basis a mixed approach could be employed.
- Please see below recommendations for each of the situations:

How to learn? Self directed approach

- We suggest the following sequence:
- Read the slides, watch the videos and reflect on the content.
- Read the additional text available where information is more detailed.
- If possible discuss with a colleague or register for a “Forum” and express your views.
- Test your understanding by taking the test at the end of the PowerPoint Presentation. Aim to score at least 70%
- If there are aspects that you find interesting and want to explore further access one of the texts recommended in the Bibliography

How to learn? Full time approach

- Attend all classes and the lecturer will provide explanations while showing the slides.
- Read the additional text available where information is more detailed.
- The lecturer will organise opportunities for you to discuss the content of the course so that you understand.
- Test your understanding by taking the test at the end of the PowerPoint Presentation. Aim to score at least 70%
- The lecturer will present a task for you to complete that will enable you to better understand the topic. The nature of the task depends on particular circumstances.
- If there are aspects that you find interesting and want to explore further access one of the texts recommended in the Bibliography

How to learn? Mixed approach

- Read the additional text available before attending the face to face classes.
- Attend all classes and the lecturer will provide explanations while showing the slides.
- The lecturer will organise opportunities for you to discuss the content of the course.
- Test your understanding by taking the test at the end of the PowerPoint Presentation. Aim to score at least 70%
- The lecturer will present a task for you to complete that will enable you to better understand the topic. The nature of the task depends on particular circumstances.
- If there are aspects that you find interesting and want to explore further access one of the texts recommended in the Bibliography

Why study Information & Research?

- To learn the basic concepts and principles related to the information search and retrieval, because you will be able to identify, select and acquire bigger and better information in less time.
- To learn the basic concepts and principles related to the scientific research methodology, because you will know which research method fits with your investigation, the stages and ethics that it must follow, where to search and what to write in a paper.
- Because you will be able to research in any field of your specialization.



Erasmus+



UNIT 4: Information retrieval from the web

MODULE 10:

SUBJECT 10.1: Information Search and Retrieval



Co-funded by the
Erasmus+ Programme
of the European Union

TOPICS 4. Information retrieval from the web

- 1. The role of the Internet in scientific research**
- 2. Internet to make research easier**
- 3. Questions**

4.1. The role of the internet in scientific research

1. Advantages for the release, diffusion and access to scientific information

- Virtual universality.
- Access and publication speed.
- Ability to update information and data published on the network immediately.
- Maximization of the use of resources.
- Free and powerful marketing.
- Cheapness.
- Great ability to quantify the use of information.
- Greater employment opportunities.



4.1. The role of the internet in scientific research

2. Problems and disadvantages

- Relativity in the advantages.
- Search expectations usually frustrated.
- Different recovery capacity of search engines.
- Information and documentary standardization is almost non-existent.
- Preponderance of the English language.
- Excessive noise during information retrieval.
- To get lost in the virtual universe.
- Problems defining virtual documents.



4.1. The role of the internet in scientific research

3. The dissemination of scientific information on the Web

- The Internet has meant, in terms of the dissemination of information, the greatest change in Humanity
 - Instantaneous publication.
 - Quick update.
 - This has generated differences as well between researchers without access to it, but fortunately it tends to disappear.
- Internet may seem that the tool for gratuity, but it is a tool for business too.
 - Large companies doing so with scientific literature
- Scientific publications available on the web similar to the paper ones, but in digital format:
 - Monographs through formats such as pdf, html, epub or electronic book.
 - Electronic journals.
 - PhD theses.
 - Conference proceedings.



4.1. The role of the internet in scientific research



Google represents both advantages and disadvantages



4.2. Tools and services

- Tools and services available to researchers to carry out their task more effectively and for keep their knowledge updated.
- Tools accessible from web browsers, Windows, Mac, Android, iOS, etc.



4.2. Tools and services

1. Google Scholar

- Perhaps the most well-known of all of them because its agglutinating objective.
- It is a version of Google specialised in scientific literature.
- Access to different types of documents:
 - Papers.
 - Reports.
 - Thesis.
 - Monographs.
 - Reviews.
 - Communications in congresses, etc.
 - Also included certain web resources.



4.2. Tools and services

- Most of researchers begin their searches by using it.
- The search allows to differentiate by language in its simple version, while the advanced one provides some filtering options for the required information:
 - To contain all the searched words
 - To contain the exact phrase.
 - To contain at least one of the written words.
 - Not to contain the written words.
 - That the words appear in the whole paper or only in the title of the same.
 - That the documents are written by a certain person or people.
 - That the documents are published in a specific publication.
 - That the documents are in the range of dates indicated.



4.2. Tools and services

- It includes the option of saving results in a particular library.
- It recognizes the authorship of the documents.
- It offers the possibility of generating a researcher profile to which our publications will be incorporated
- It also presents some problems as its opacity, since it is not possible to know its algorithm, so it entails:
 - Lack of control over some of the sources included.
 - The concept of "relevance" in the ordering of search results.
 - The repetition of content.

4.2. Tools and services

2. Scopus

- Scopus is an abstracts and citation database of approximately 13,000 scientific journals.
- The quality of the journals is reviewed every year to prepare the SCImago Journal Rank.
- Scopus, which belongs to Elsevier, is not open access, so it requires a subscription to access its contents.



4.2. Tools and services

3. Web of Science

- The Web of Science (WOS) offers the references of scientific publications from different disciplines.
- It is also a private initiative, in this case owned by Thompson Reuters, so it requires a subscription to access its contents.
- It is integrated into the ISI Web of Knowledge and allows access to the texts contained in its bibliographic databases through citations and references.
- Like Scopus, it produces its own impact indexes, such as the Science Citation Index or the Social Sciences Citation Index.



4.2. Tools and services

4. Sci-Hub

- There are other resources belonging to large publishers or universities, such as JSTOR, ProQuest or EBSCO.
- Science is largely dependent on the services that these publishers offer.
- It is a situation that has been criticized, so that alternatives have emerged
 - Sci-Hub.
- Created in 2011.
- Its objective is to provide free access to the databases of these institutions through the access that the researchers whose centres have assigned.



4.2. Tools and services

- When someone uses the Sci-Hub aggregator, it performs a search in LibGen and if the paper is in that repository, it returns it.
- Otherwise, Sci-Hub use the credentials of a researcher to return it into full text to the researcher who has no access.
- It is illegal in some countries.
- It is an alternative for those researchers without access to these digital objects.

4.2. Tools and services

5. Scholarly collaboration networks

- Legal alternatives for free-access:
- ResearchGate
 - Social network of scientists
 - Search engine for its own resources and others such as PubMed or arXiv.
 - It allows to follow other researchers and the creation of working groups.
- Academia.edu
 - Similar characteristics and purposes that ResearchGate.
 - It does not guarantee open access to collections.
- Mendeley
 - Tool that has generated a social network over the years,
 - Another main objective.



4.2. Tools and services



ResearchGate



4.2. Tools and services

6. Academic identifications

- A common problem to all tools is to correct indexing and acknowledging responsibility.
- Several initiatives have arisen to control the authorship in a universal way.
- There are identifiers granted by international recognition and which the journals themselves request to authors to publish their papers.
- The most recognized is the Open Researcher and Contributor ID (ORCID).
- Researcher ID (Thompson Reuters)
- Scopus Author ID (Elsevier).



4.2. Tools and services

7. Bibliography management

- A bibliographic manager is a tool that allows the storage of bibliographic references in an organized and controlled manner.
- Mendeley was created as a bibliographic manager.
 - It can be used from the web and at the same time has Windows, Mac, Linux, Android and iOS applications.
 - It allows the creation of bibliographies and citations in any style.
 - It also lets to import and export these bibliographies in different extensions.

4.2. Tools and services

- Zotero is a similar tool that also has a web and a standalone version. They have some other functions useful for researchers:
 - Creation of bibliographies.
 - Possibility of creating different folders in which to organize documents.
 - Automatic import and update metadata of documents.
 - Edition of document metadata.
- In addition to the Google Scholar, Mendeley and Zotero, there are other initiatives related to bibliographic management:
 - BibTeX.
 - EndNote.
 - RefMan.
 - RefWorks.
 - NoodleTools.
 - EasyBib.
 - RefDot.
 - Bibme.
 - Citation Machine
 - Citelighter.



5.3. Questions

- 1) Make 5 different searches in Google Scholar using the same words but using each time one different characteristic of the advanced search (or Boolean operators). Compare the results
- 2) Does the Internet bring more advantages or disadvantages to scientific research? Why?
- 3) Create 2 different profiles or ID and search for other researchers of your institution through 2 or more of the tools abovementioned. Compare the results.

Key references

- Bartling, S., & Friesike, S. (2014). *Opening science: The evolving guide on how the internet is changing research, collaboration and scholarly publishing*. Springer-Verlag GmbH.
- Büttcher, S., Clarke, C. L., & Cormack, G. V. (2016). *Information retrieval: Implementing and evaluating search engines*. Mit Press.
- Chaín Navarro, C. (2004). *Técnicas y métodos de recuperación de la información*. Diego Marín
- Cooper, L. B., & Perez-Martinez, M. (2015). Discovery & Born-Digital Archiving: Open Source Systems for Preservation and Access.
- Gerakopoulou, E., Founti, C. M., & Foster, C. (2018). Removing all barriers in the way of Science: A case study on Sci-Hub's usage in Europe. *iConference 2018 Proceedings*.
- Gil, L. (2015). *Google Scholar: el buscador académico con mayor impacto*. Social Media en Investigación.
- Harzing, A. W., & Alakangas, S. (2016). Google Scholar, Scopus and the Web of Science: a longitudinal and cross-disciplinary comparison. *Scientometrics*, 106(2), 787-804.
- Roy, B. K. (2015). *Institutional Digital Repositories: From Policy to Practice*. LAP LAMBERT Academic Publishing.
- Van Noorden, R. (2014). Online collaboration: Scientists and the social network. *Nature news*, 512(7513), 126.

Module 10 Information & Research

Learning Unit 5 – Databases and Research

Authors

Alberto Hoces-García – Universidad de Murcia

Celia Chaín-Navarro – Universidad de Murcia

Licences

- This training material is subject to Creative Commons licenses. For training material, such as images, subject to another type of license, the license is explicitly stated
- This material is licensed under **Creative Commons Attribution-NonCommercial-ShareAlike 4.0 International License (CC BY-NC-SA 4.0)**.

To see a copy of this license, please visit

<https://creativecommons.org/licenses/by-nc-sa/4.0/deed.en>

Funding

- This training material has been developed within the framework of the **International Master for the Furniture Sector (IM-FUTURE)** project.
- The project is implemented within the framework of the **Cooperation for Innovation and the Exchange of Good Practices** Key Action and is funded by the **Erasmus+ Programme** of the **European Union**.

Learning outcomes

- This module will enable you to:
Understand the importance of materials and the connection between materials' properties and engineering technologies required to process them.
- Understand and critically evaluate aspects related to production organisation and optimisation
- Understand and critically evaluate aspects related to automation in furniture manufacturing
- Ability to understand and critically analyse the eco-sustainability aspects in furniture production

How to learn?

- This course could be taken using a “self directed” learning approach.
- For those attending “full time” courses a lecturer will be directing your learning using the learning materials available.
- For those enrolled on a “part time” basis a mixed approach could be employed.
- Please see below recommendations for each of the situations:

How to learn? Self directed approach

- We suggest the following sequence:
- Read the slides, watch the videos and reflect on the content.
- Read the additional text available where information is more detailed.
- If possible discuss with a colleague or register for a “Forum” and express your views.
- Test your understanding by taking the test at the end of the PowerPoint Presentation. Aim to score at least 70%
- If there are aspects that you find interesting and want to explore further access one of the texts recommended in the Bibliography

How to learn? Full time approach

- Attend all classes and the lecturer will provide explanations while showing the slides.
- Read the additional text available where information is more detailed.
- The lecturer will organise opportunities for you to discuss the content of the course so that you understand.
- Test your understanding by taking the test at the end of the PowerPoint Presentation. Aim to score at least 70%
- The lecturer will present a task for you to complete that will enable you to better understand the topic. The nature of the task depends on particular circumstances.
- If there are aspects that you find interesting and want to explore further access one of the texts recommended in the Bibliography

How to learn? Mixed approach

- Read the additional text available before attending the face to face classes.
- Attend all classes and the lecturer will provide explanations while showing the slides.
- The lecturer will organise opportunities for you to discuss the content of the course.
- Test your understanding by taking the test at the end of the PowerPoint Presentation. Aim to score at least 70%
- The lecturer will present a task for you to complete that will enable you to better understand the topic. The nature of the task depends on particular circumstances.
- If there are aspects that you find interesting and want to explore further access one of the texts recommended in the Bibliography

Why study Information & Research?

- To learn the basic concepts and principles related to the information search and retrieval, because you will be able to identify, select and acquire bigger and better information in less time.
- To learn the basic concepts and principles related to the scientific research methodology, because you will know which research method fits with your investigation, the stages and ethics that it must follow, where to search and what to write in a paper.
- Because you will be able to research in any field of your specialization.



Erasmus+



UNIT 5: Databases & research

MODULE 10:

SUBJECT 10.1: Information Search and Retrieval



Co-funded by the
Erasmus+ Programme
of the European Union

TOPICS 5. Scientific publication

1. Impact indexes
2. Digital repositories
3. Aggregators
4. Questions

5.1. Impact indexes

- The use of journals is now the standard of practically all the areas of the Science.
- Impact indexes are adapted as far as possible to almost every area from the qualitative science.
- Already entered the 21st century, national and international impact indices began to be used regularly for those who wanted to improve their curriculum.
- Most of the university libraries offer information on these tools on their websites, specifying their characteristics.

5.1. Impact indexes

- These ways of measuring the quality of publications based on their impact are not uniform, and there are several instruments for this.
- Some are only directories, which without providing an impact figure, analyse the publications using a series of criteria and include those that they consider to be of higher quality:
 - ERIH +, Art and Humanities Citation Index
 - FECYT quality seal.
- Others provide a figure that position the journal within a ranking:
 - Social Science Citation Index.
 - SCImago.
 - Journal Scholar Metrics.
 - Latindex.
 - CARHUSPlus.
 - CIRC.



5.1. Impact indexes

- Each listing contains multiple publications, but they are not always the same, nor are they in the same positions.
 - Important journals are missing sometimes.
- The Journal Citation Report (JCR) is the most valued international list.
- And SCOPUS, which is its competence in Europe and Latin America.

5.2. Digital repositories

- Those that contain the complete texts, although we focus on those that offer relevant information for research and innovation.
- A permanent repository of documents in digital version of an institution that are accessible free of charge and in full text for all citizens from Internet.

5.2. Digital repositories

- Their main features are:
 - Their intention of permanence.
 - Institutions are "owners" of these digital documents.
 - They allow self-archiving.
 - The inclusion of documents is voluntary.
 - They use common international standards.
 - They use a specific software with free access.
 - They offer the option of a basic type of search and an advanced one.
 - They allow the user to search in different ways according to the contents included.
 - They are very similar, easy to use and free.
 - They contain links to other repositories, collectors and directories.
 - They comply with the international regulations issued in this regard.
 - They allow the quantification of the archived information.
 - They contain a huge and heterogeneous number of documents from multiple disciplines, full text.



5.2. Digital repositories



RIJKS STUDIO

Discover the possibilities of 125,000 masterpieces

Powered by BankGiroLoterij

Start your own Rijksstudio >

Rijksstudio is de digital repository of the
Rijksmuseum of Netherlands



5.2. Digital repositories

- One of the best lists of digital repositories is Webometrics. In its ranking, the first ones that appear are four American repositories: Social Sciences, Experimental Sciences of Cornell, a system of astrophysical data Smithsonian / NASA and CiteSeerX.
- The option of free access multiplies the visibility, so facilitates the impact. As researchers, we are interested in having our works consulted and cited.
 - An article published in a digital repository (or in an electronic journal contained there) has between 4 and 10 times more chance of being read, and therefore cited.



5.2. Digital repositories



**Digital repositories made possible
worldwide access**

5.3. Aggregators

- The difference between a digital repository and an aggregator is that the first has the "property" of the documents.
- The control of the file represents the capacity to offer it over time.
- Collectors / aggregators only direct to the file and indicate which institution has it.
- The great advantages of aggregators are that they allow access to a greater number of documents and that they are much quicker to search for information.
- But they cannot guarantee the permanence of the documents to which they give access.

5.3. Aggregators



Europeana is the European Union aggregator which aims to collect all the digital collections of European institutions



4.4. Questions

- 1) Make a list of 10 journals related to the furniture level at least with at least 5 of them of high impact.
- 2) Are there any repositories related to the furniture industry? Describe them.
- 3) Are there any aggregators related to the furniture industry? Describe them.

Key references

- Bartling, S., & Friesike, S. (2014). *Opening science: The evolving guide on how the internet is changing research, collaboration and scholarly publishing*. Springer-Verlag GmbH.
- Buela-Casal, G. (2010). Scientific journal impact indexes and indicators for measuring researchers' performance. *Revista de Psicodidáctica*, 15(1).
- Büttcher, S., Clarke, C. L., & Cormack, G. V. (2016). *Information retrieval: Implementing and evaluating search engines*. Mit Press.
- Chaín Navarro, C. (2004). *Técnicas y métodos de recuperación de la información*. Diego Marín
- Cooper, L. B., & Perez-Martinez, M. (2015). *Discovery & Born-Digital Archiving: Open Source Systems for Preservation and Access*.
- Roy, B. K. (2015). *Institutional Digital Repositories: From Policy to Practice*. LAP LAMBERT Academic Publishing.