

Subject Analysis, Content Analysis, and Domain Analysis: Concepts, Methods, and Applications

Tauany Lorena Alves Silva Portella  and Gercina Ângela de Lima 
Federal University of Minas Gerais (UFMG), Belo Horizonte, Brazil

From the perspective of the constant increase in data and information, consider in Library and Information Science that the correct analysis and representation of the contents of documents analyzed in specific domains is essential for the retrieval, organization, and dissemination of information. Subject Analysis categorizes topics and details, making it easier to retrieve relevant information. Domain Analysis studies specific characteristics of a field of knowledge, comprising terminologies and concepts. Content Analysis identifies and analyzes textual elements, deepening the understanding of documentary content. This study explores these analyses' approaches, techniques, and methodologies, highlighting their often confused interrelationships, differences, and similarities. To achieve the proposed objective to support the conceptual and theoretical-methodological discussion on subject analysis, content analysis, and domain analysis, focusing on their interrelations, differences, and similarities that are often confused in their concepts and methodologies, the research developed an exploratory and descriptive approach, a bibliographic survey was carried out in the BRAPCI database, using the terms "domain analysis", "content analysis" and "content analysis", recovering 134 documents. Results are efficiently defined and applied to each analysis. These analyses guarantee efficient information retrieval, which is vital to growing data volume.

Keywords: subject analysis, domain analysis, content analysis, semantic analysis

Introduction

Given the constant increase in data and information, the correct analysis and representation of document contents and a specific domain are of the utmost importance for retrieving and disseminating knowledge. Among the main available approaches are subject analysis, domain analysis, and content analysis, techniques used to identify, interpret, and understand the meaning of document content and knowledge domain.

Subject analysis aims to identify the main subjects in the text, extracting these concepts to be used in search and facilitating information retrieval. It allows the identification of dominant terms for use in classification. Domain analysis also seeks to identify themes within the text, but it considers the domain in which it is addressed, considering terminologies, concepts, and specific relationships within an area.

With a broader approach, content analysis seeks to understand implicit meanings in texts. It can analyze qualitative and quantitative data and the relationships between terms to arrive at a more detailed understanding of document content.

In this article, we will explore and compare these three

text analysis approaches—subject analysis, domain analysis, and content analysis—examining their characteristics, methods, and applications. Additionally, we will discuss their advantages and limitations and how these techniques can be combined to achieve a more comprehensive and in-depth understanding of textual content.

Methodology

This research has an exploratory and descriptive nature to achieve the proposed objective, starting with a literature review, aiming to analyze and describe the results. Thus, bibliographic research was conducted to support the conceptual and theoretical-methodological discussion on subject, content, and domain analysis, focusing on their interrelationships, differences, and similarities, which are often confusing in their concepts and methodologies.

It justifies itself by highlighting the need for more awareness regarding subject analysis in Information Science, which can lead to confusion regarding its concept, methodology, and purpose compared to content analysis and domain analysis.

The models revealed differences in the type of strategy used in text treatment, the approach type, and the theoretical stance in which they fit.

The Definitions

Subject Analysis (SA)

Subject analysis (SA) emerged as a need to organize information in documents, making their retrieval easier. The first bibliographic classification systems appeared in the 19th century when the quantity of library books grew significantly. Since then, the concern of finding efficient ways to organize information in documents has become increasingly prominent (Weitzel, 2002).

Subject analysis is a technique that originated in library science and information science, aiming to assist in the organization and retrieving information in documents. It dates back to the development of bibliographic classification systems in the 19th century to organize library books systematically. Over time, subject analysis has evolved to include more sophisticated techniques for information organization and retrieval, such as using keywords for information retrieval.

Cesariano and Pinto (1980) emphasize that subject analysis is the foundational operation for information retrieval. Naves (2000, p. 249) provides:

- a definition very close to theirs,
- describing subject analysis as a foundational operation in subject indexing,
- encompassing the process through which the indexer extracts the content of a document.

Lima (2020, p. 9) points out that subject analysis is an intellectual activity, therefore subjective, starting with the technical reading of the main parts of a document. It requires the indexer's linguistic, cognitive, and logical knowledge to determine the document's subject. Thus, as the author states, it involves subjective procedures due to its intellectual nature.

Reading to understand and identify the text is essential to later selecting valid terms to represent the content of a document (LIMA, 2020, p. 9).

Laville and Dione (1999) also consider subject analysis a primarily intellectual and subjective activity. In this activity, all acquired knowledge and experiences are sharpened and reflected in the indexer's interpretation, aiming to objectify subjectivity.

In Library Science and Information Science, subject analysis is the first stage of the indexing process, which involves two steps (analysis and translation). However, in the literature, there are suggestions of three or even four stages in its execution. Nevertheless, the UNISIST proposal (1981) with three stages is the most cited in the literature: 1) understanding the content of the document as a whole, (2) identifying concepts representing this content, and (3) selecting concepts for retrieval (UNISIST, 1981).

It is important to note that, in this process, the adopted methodological parameters should always consider the context in which the documents were produced and assess the relevance of the content for a specific audience or objective.

Content Analysis (CA)

Content analysis (CA) is a technique used in Information Science to understand and interpret qualitative data. It involves the systematic analysis of the content of a set of data, such as interviews, documents, and other sources of information.

In Information Science, content analysis is often employed in user studies and research aimed at understanding the needs and behaviours of library and information system users. The technique can be applied to different data types, including textual, visual, and auditory. Content analysis involves identifying themes and patterns in the data and interpreting and explaining these themes. The technique can be carried out manually or with the assistance of specific software tools.

According to Cunha (1983), this research technique is much older than its denomination suggests. He points out that studies investigating mental content, personal documents, or interview reproductions have already used similar methodologies. For the author, content analysis addresses questions relevant to the analyzed object. In communication, he highlights (as cited in Berelson, 1971, p. 2) content analysis as a "research technique that aims for the objective, systematic, and quantitative description of the manifest content of communication." However, it is interesting to note that when he turns to the field of Social Sciences, more specifically Library Science, he refers to it as Subject Analysis. In his methodology, he categorizes procedures to initiate the analysis: the unit of analysis, defined as "individual units on which descriptive and explanatory statements should be made" (Cunha, 1983, p. 249). Another procedure is "Sampling," where a sample of what will be studied is selected, similar to fragmenting a document and choosing a part for analysis. Additionally, there is the coding procedure, referring to quantifying qualitative research data for reproduction purposes. The definition and meaning of codes should be relevant for better research replication.

Câmara (2013) tells us that

in the 1920s, after World War I, due to Leavell's studies on propaganda employed in that war, content analysis gained systematic forces of use, taking on organized forms of investigative methods. World War II intensified the development of propaganda, and therein, it gained importance, allowing many disciplines, including linguistics and mainly autobiographical literature, to show a preference for its use. Psychoanalysis and clinical psychology use it as one of the elements for interpreting the individual's life. (p. 159)

Bardin (1977) proposes a methodology for content analysis that consists of three phases:

- **Pre-analysis:** This phase involves an exploratory approach to the material to be analyzed. The goal is to identify the material's characteristics, such as its nature, origin, form, and content. It is also essential to define the purpose of the analysis, the criteria for selecting data, and the categories of analysis.
- **Analytical description (Exploration of the material):** In this phase, the material is systematically analyzed based on the categories defined in the pre-analysis. The goal is to describe the material's content objectively and precisely, identifying the text's main characteristics and elements.
- **Interpretative inference (Handling the obtained results and interpretation):** In this phase, the analysis becomes more explanatory based on the inferences made in the analytical description. The goal is to interpret the content, identifying patterns, trends, relationships, and underlying meanings.

The methodology proposed by Bardin for content analysis involves a systematic and rigorous approach, allowing for a deeper understanding of the content of the material being analyzed. Bardin emphasizes that the history of content analysis is marked by constant evolution, with new approaches and techniques being developed over time to deal with the complexity and diversity of the analyzed data. It is intriguing how she highlights that 'Content analysis can be an analysis of "meanings" (e.g., thematic analysis), although it can also be an analysis of "signifiers" (lexical analysis, analysis of procedures.)' (Bardin, 1977, p. 34).

Interestingly, Bardin mentions that content analysis is an analysis of signifiers because it brings us back to subject analysis, where she later states that the descriptive treatment would not be exclusive as the starting point solely for content analysis.

Domain Analysis (DA)

Domain analysis (DA) is investigating and understanding a specific thematic domain's contents, concepts, and terminology. According to various authors, the definition of domain analysis may vary depending on the adopted theoretical perspective but generally involves selecting and organizing concepts and terms related to a specific theme.

According to Hjørland & Albrechtsen (1995), domain analysis is

a theoretical approach to Information Science (IS) states that the best way to understand information in information science is to study areas

of knowledge as "discourse communities, which are parts of the division of labor society. Organization of knowledge, structure, patterns of cooperation, language and forms of communication, information systems, and relevance criteria are reflections of the objects of the work of these communities and their role in society. (Hjørland & Albrechtsen, 1995, p.400).

Besides, Hjørland (2002) states that domain analysis seeks to map and understand the characteristics and properties of a particular domain, providing a solid foundation for the organization and access to information. Domain analysis is a research technique that identifies and describes a specific knowledge field's concepts, themes, and sub-themes. In a theoretical approach to Information Science (IS), the best way to understand the information is to study knowledge areas as "communities of discourse," which are parts of the division of labour in society.

Furthermore, Hjørland and Albrechtsen (1995) propose that domain analysis should be understood as an investigation that encompasses not only content and terminologies but also social structures and discursive practices within a domain. Domain analysis is not just about 'investigating and understanding the contents, concepts, and terminology of a specific thematic domain.' As Hjørland (2002) and Smiraglia (2015) point out, it is a comprehensive study encompassing the dynamics and interactions between concepts, practices, and users within a given context. Smiraglia (2015) expands this view by exploring how domains are shaped by the interactions between their components and users. Additional studies, such as those by Tennis (2003) and Beghtol (2003), also offer valuable perspectives on the methodologies and applications of domain analysis, highlighting the importance of considering both professional classifications and more intuitive ones.

Regarding domain analysis, it is essential to consider the article by Hjørland (2017) published in the Encyclopedia of Knowledge Organization, where the author substantially updates his 2002 article. In this update, he expands this perspective, introducing new approaches that consider the more complex dynamics and interactions within domains, highlighting the importance of understanding the social and cultural contexts that influence the formation and evolution of knowledge domains. Furthermore, it incorporates emerging methodologies that use advanced information technologies, such as network analysis and data mining, to map and analyze domains more comprehensively and in a detailed way.

It is a technique that originated in different areas of knowledge, including Information Systems and Library Science. According to Hjørland and Albrechtsen (1995), in Library Science, domain analysis originates in the study of classification and cataloguing, aiming to understand the characteristics of the objects and concepts that will be organized.

In Information Systems, domain analysis aims to understand the study domain of an information system and identify the specific requirements of the system. According to Pressman (2016), domain analysis is one of the most critical activities in the software engineering process, as it allows developers to understand the requirements and needs of users better.

Despite emerging in different areas, domain analysis in Information Science has the common goal of understanding the study domain and identifying the specific requirements of a particular system or object of study. Over time, various methodologies and approaches have been developed to conduct domain analysis in different areas of knowledge.

Moreover, Domain analysis is a fundamental process for constructing knowledge organization systems, as it identifies the most relevant terms and concepts in a given thematic domain. For these authors, it is a dynamic and continuous process that should be constantly updated to keep up with changes.

The methodology proposed by Hjørland (2002) for domain analysis in information science involves the following steps:

- **Domain Identification:** Define the domain scope to be analyzed.
- **Identification of Key Theories, Concepts, and Methods:** Identify the fundamental theories, concepts, and methods used in the domain area and identify relevant information sources.
- **Identify the primary information sources relevant to the domain,** including books, articles, and databases.
- **Critical Analysis of Information Sources:** Conduct a critical analysis of identified information sources, considering different theoretical and methodological perspectives.
- **Identification of Key Documents:** Identify the key documents relevant to the domain, including reports, standards, and laws.
- **Critical Analysis of Key Documents:** Critically analyze the identified key documents, considering different theoretical and methodological perspectives in the field.
- **Identification of Key Actors:** Identify the critical actors involved in the domain, including researchers, professionals, and institutions.
- **Critical Analysis of Key Actors:** Conduct a critical analysis of the identified vital actors, considering different theoretical and methodological perspectives in the field.
- **Identification of Gaps in the Domain:** Identify gaps in the domain that need to be addressed by new research and studies.

- **Synthesis of Results:** Synthesize the domain analysis results into a report presenting critical theories, concepts, methods, relevant information sources, key documents, actors, and identified gaps in the domain.

To elucidate the characteristics and similarities among the analyses described above, TABLE 1 presents the definitions of subject, content, and domain analysis based on the literature, as proposed by authors who study and apply these techniques.

The Interfaces Between Techniques of SA, CA, and DA

As mentioned, SA, CA, and DA techniques identify, interpret, and understand the meaning of document content and knowledge domains. This work aimed to explore the interfaces between these three analysis approaches to elucidate their characteristics, methods, applications, and relationships.

Subject Analysis and Content Analysis

Subject analysis and content analysis are two research techniques that are closely related. Both aim to analyze and understand the content of a text, but they differ in some aspects. Subject analysis focuses on identifying the main themes addressed in a document to assist in organizing and retrieving information in a collection or database. Subject analysis can be done manually by reading and identifying the addressed themes or using specialized software.

On the other hand, content analysis is a technique aimed at understanding the meaning of the text as a whole. It is primarily used in qualitative research, where the goal is to understand individuals' perceptions and opinions regarding a specific topic. Content analysis involves a systematic and rigorous approach to better understanding the analyzed material.

Despite different objectives, subject and content analysis can be used together to comprehensively analyze a document's content. For example, subject analysis can identify the main subjects addressed in a text and content analysis can be employed to understand the meaning of these themes and how they relate to the broader research context.

Subject analysis and content analysis are two research techniques that complement each other and can be used together to more thoroughly analyze a text's content. Bardin (1977) considers subject analysis one of the stages of content analysis.

Subject Analysis and Domain Analysis

Subject analysis and domain analysis are related techniques but have different objectives. While subject analysis aims to identify the themes present in documents, domain analysis aims to identify the specific concepts and terms of a particular field of knowledge.

Domain analysis is also used in areas beyond Information Science to assist in organizing information in specialized documents and software development. For example, in a medical library, domain analysis would be used to identify concepts and specific terms in the medical field, such as diseases, treatments, and medications. Another application of domain analysis is understanding user and business needs for the software's application.

On the other hand, subject analysis would be used to represent the specific subjects in documents, regardless of the knowledge domain covered. For instance, in a library with books from various knowledge areas like history, literature, and sciences, subject analysis is used to identify the subjects covered in the books and subsequently represent them as terms such as "history of art," "classic novels," and "marine animals."

Hjørland (2004) emphasizes that in domain analysis, subject analysis is used to identify the main topics and concepts present in the domain, creating a controlled vocabulary and defining terms and concepts used in the area. To illustrate the interfaces between the analysis techniques, a comparative table, TABLE 2, is presented. Structured based on eight elements considered necessary for highlighting this comparison, namely: (1) definition, (2) characteristics, (3) objectives, (4) functions, (5) procedures, (6) approach, (7) originating area, and (8) key authors.

As observed, we aimed to include **definitions** presenting each technique's main characteristic in the definition element. While subject analysis appears as a technique linked to representing the content of a document, content analysis is considered a set of communication analysis techniques, and domain analysis has a more theoretical approach with the objective of understanding and studying an area of knowledge.

Regarding the **characteristic** element, it is evident that all are considered analysis techniques, although sometimes they are also considered by some authors as methods or even.

Regarding the **objectives** element, it is clear that while subject analysis aims to condense a document's information, content analysis seeks to interpret the information to understand the meanings of a particular theme, and domain analysis focuses on analyzing the coherence of a knowledge domain.

Subject analysis represents retrieval; content analysis qualitatively and quantitatively analyzes the occurrences of specific terms within a theme. Domain analysis seeks to understand a domain in a specific context.

Regarding **procedures**, the subject analysis relies on technical reading to understand, identify, and select valid concepts for representation aiming at retrieval. Content analysis is carried out in three stages, with the first stage, pre-analysis, having the most detailed procedures. Domain analysis presents itself with ten procedures suggested by Hjørland (2002).

The subject analysis is characterized as qualitative because it is considered a subjective process, while the content analysis

approach, besides being qualitative, has an extreme qualitative bias. In domain analysis, the approach is theoretical and methodological.

Subject analysis **originated** in Library Science, while the development of content analysis occurred in linguistics, mainly with the studies of Bardin (1977), which contributed to popularizing the technique. Despite domain analysis having influences from different areas, its initial studies occurred in computer science and later in information science with the studies of Hjørland and Albrechtsen (1995).

The **principal authors** who supported these studies, from the perspective of this article, were: in subject analysis (Foskett (1973), Lancaster (2004), Dias Naves (2007), Fujita (2006, 2007, 2008, 2009, 2010); Lima (2020, 2021) ABNT NBR 12676 (1992); in content analysis (H. Lasswell (2013), Laurence Bardin (1977); and in domain analysis (Hjørland & Albrechtsen (1995), Hjørland (2002, 2004), Tennis (2003), Smiraglia (2012).

Final Considerations

Subject, domain, and content analyses are essential for understanding and organizing information. Subject analysis identifies the main topics of a document, domain analysis understands the thematic area in which it is inserted, and content analysis explores meanings and comprehends relationships within the document's intrinsic information.

These analyses provide procedures that allow for the efficient study of informational and terminological content based on analyzing an area's content, theme, and domain, enabling the representation and retrieval of information. Moreover, they assist in discovering relevant information by decoding what is implied, proving crucial for the organization and preservation of information.

Based on the results of this research it is expected to clarify the concepts and methodologies of subject analysis, content analysis, and domain analysis, as well as understand their relationships to contribute to understanding these concepts in information science.

References

- Bardin, L. (1977). *Análise de conteúdo*. Lisboa: Edições 70.
- Beghtol, C. (2003). Classification for information retrieval and classification for knowledge discovery: relationships between "professional" and "naive" classifications. *Knowledge organization*, 30(2), 64-73.
- Berelson, B. (1971). *Content analysis in communication research*. Hafner.
- Câmara, R. H. (2013). Análise de conteúdo: da teoria à prática em pesquisas sociais aplicadas às organizações. *Gerais: Revista Interinstitucional de Psicologia*, 6(2), 179-191.
- Cesarino, M. A. N., & Pinto, M. C. M. F (1980). Análise do assunto. *Revista de biblioteconomia de Brasília*, 8(1).

- Da Cunha, M. B. (1983). Análise de conteúdo: uma técnica de pesquisa. *Revista de biblioteconomia de Brasília*, 11(2), 247-256.
- Da Rocha Weitzel, S. (2002). O desenvolvimento de coleções e a organização do conhecimento: suas origens e desafios. *Perspectivas em Ciência da Informação*, 7(1).
- De Oliveira, M. L., & Carosia, J. S. (2022). Modelagem de um sistema de controle de portaria eletrônica remota para uma empresa de monitoramento eletrônico. *Congresso de Tecnologia-Fatec Mococa*.
- Dias, E. W., & Naves, M. M. L. (2007). *Análise de assunto: teoria e prática*. Thesaurus.
- Foskett, A. C. (1973). *A abordagem temática da informação*. Polígono.
- Fujita, M. S. L. e Rubi, M. P. (2006). O ensino de procedimentos de política de indexação na perspectiva do conhecimento organizacional: uma proposta de programa para a educação à distância do bibliotecário. *Perspectivas em ciência da informação*, 11, 48-67.
- Fujita, M. S. L. (2007). *O contexto da leitura documentária de indexadores de bibliotecas universitárias em perspectiva sociocognitiva para a investigação de estratégias de ensino*. UNESP.
- Fujita, M. S. L., Ferreira, G. I. S., & de Oliveira Inácio, M. (2008). Ensino do processo de análise de assunto para indexação com aplicação de um modelo de leitura: estudo de avaliação comparada em cursos de graduação em Biblioteconomia do Brasil. *Ibersid: revista de sistemas de información y documentación*, 2, 163-176.
- Fujita, M. S. L., Rubi, M. P., & Boccato, V. R. C. (2009). O contexto sociocognitivo do catalogador em bibliotecas universitárias: perspectivas para uma política de tratamento da informação documentária. *DataGramaZero: Revista de Ciência da Informação*, 10(2).
- Fujita, M. S. L. (2010). O contexto profissional do indexador no ensino de indexação. *Encontros Bibli: revista eletrônica de biblioteconomia e ciência da informação*, 15(30), 91-104.
- Hjørland, B. (2002). Domain analysis in information science: eleven approaches—traditional as well as innovative. *Journal of documentation*, 58(4), 422-462.
- Hjørland, B., & Albrechtsen, H. (1995). Toward a new horizon in information science: Domain-analysis. *Journal of the American Society for Information Science*, 46(6), 400-425.
- Hjørland, B. (2017). Domain analysis. *Knowledge organization*, 44(6), 436-464.
- Lancaster, FW (2004). *Indexação e resumos*. Tradução de AA Briquet de Lemos.
- Laville, C., & Dionne, J. (1999). A construção do saber (Vol. 340). Belo Horizonte: UFMG.
- Lasswell, H. D. (2013). *Propaganda Technique In The World War*. Martino Fine Books.
- Lima, G. Â. B. de O., & Maculan, B. C. M. dos S. (2014). Análise de assunto a partir de uma perspectiva histórica do ARIST. *Ciência da Informação*, 43(1).
- Lima, G. Â. de. (2020). O ensino da Análise de Assunto: Em busca de uma metodologia. *Informação & Sociedade*, 30(4), Article 4. <https://doi.org/10.22478/ufpb.1809-4783.2020v30n4.57182>
- Ranganathan, S. R. (1967). *Prolegômenos para classificação de bibliotecas*. Asia Publishing House.
- Smiraglia, R. (2015). *Domain analysis for knowledge organization: tools for ontology extraction*. Chandos Publishing.
- Silva, C., Gobbi, B. C., & Simão, A. A. (2005). O uso da análise de conteúdo como uma ferramenta para a pesquisa qualitativa: descrição e aplicação do método. *Organizações Rurais & Agroindustriais*, 7(1R.), 70-81.
- Tennis, J. (2003). Two axes of domains for domain analysis. *Knowledge organization*, 30(3/4), 191-195.
- Trivisios, A. N. (1987). Introdução à pesquisa em ciências sociais. *A pesquisa*, 133.
- Da Cunha, M. B. (1983). Análise de conteúdo: uma técnica de pesquisa. *Revista de Biblioteconomia de Brasília*, 11(2), 247-256.
- UNISIST (1981). Princípios de indexação. *Revista da Escola de Biblioteconomia da UFMG, Belo Horizonte*, 10(1), 83-94.

Appendix A: Interfaces between SA, CA, and DA

Subject Analysis Definitions		Content Analysis Definitions		Domain Analysis Definitions	
Berelson (1952)	Subject analysis is a research technique to identify and categorize the main themes addressed in a text based on predefined categories.	Berelson (1952)	Content analysis is a research technique that aims to objectively, systematically, and quantitatively describe the manifested content of communication.	Neighbors (1984)	An attempt to identify objects, operations, and relationships among those that the expert perceives to be important in the domain.
Bloom (1956)	Subject analysis is a research technique that identifies the themes and subthemes a text addresses and categorizes them into predefined categories.	Weber (1990)	Content analysis is a research technique for discovering and describing the characteristics of a text's content, such as themes, concepts, values, and attitudes.	Rubén Prieto-Díaz (1990)	Domain analysis is a process by which information used in the development of software systems is identified, captured and organized with the purpose of making it reusable in the creation of new systems.
Taylor (1968)	Subject analysis is defined as identifying and describing the main themes present in a document, intending to enable its subsequent retrieval.	Bardin (1994)	Content analysis is now defined as a set of techniques for analyzing communications, utilizing systematic and objective procedures for describing the content of messages and indicators (quantitative or not) that allow the inference of knowledge regarding these messages' production/reception conditions (inferred variables).	Hjørland & Albrechtsen, (1995).	It is a theoretical approach to Information Science (IS), which states that the best way to understand information in information science is to study areas of knowledge as —discourse communities, which are parts of society's division of labor. Knowledge organization, structure, cooperation patterns, language and forms of communication, information systems and relevance criteria are reflections of the objects of work of these communities and their role in society. Psychology, knowledge, the need for information and subjective relevance criteria must be seen from this perspective.
Foskett (1973, p. 40)	The key operation of indexing, deciding what the document is, remains less discussed and less reducible to rules.	Miles; Huberman (1994)	Content analysis is searching for systematic patterns of meaning in text data.	Hjørland (1998; 2002)	Domain analysis is an activity that seeks to map and understand the characteristics and properties of a given domain, in order to provide a solid basis for the organization and access to information.
Cesarino e Pinto (1980)	Base operation for the entire information retrieval procedure occurs in two situations: upon entry of the document into the system and when there is a request for information.	Godoy (1995)	Content analysis is a methodological technique that can be applied to various discourses and all forms of communication, regardless of the nature of their medium. In this analysis, the researcher seeks to understand the characteristics, structures, or patterns that underlie the fragments of messages under consideration. The analyst's effort is twofold: to understand the meaning of communication as if being the average recipient and, primarily, to shift the focus, seeking another meaning, another message, visible through or alongside the first.	Nielsen (2000, p. 10).	Domain analysis is seen as the natural starting point of an iterative process of building a thesaurus consisting of several interactive subprocesses.
Lancaster (1986)	Subject analysis is a process of identifying and selecting terms or concepts that represent the content of a document.	Freitas, Cunha, Moscarola (1997)	Content analysis, currently, can be defined as a set of methodological tools, in constant refinement, that are suited for analyzing different sources of content (verbal or non-verbal).	Motta e Vasconcellos (2006)	Domain analysis is a research technique that allows you to identify and describe the concepts, themes and subthemes that make up a given field of knowledge.
Lancaster, Elliker, Connell, (1989)	Presence, identification, and expression of subject matter in textual documents, databases, controlled vocabularies, natural languages, information requests, and search strategies.	Graneheim; Lundman (2004)	Content analysis is a research technique that aims to identify and describe the main characteristics and elements of a text's content based on a set of predefined categories.	Santos e Palva (2009)	Domain analysis is a fundamental process for building knowledge organization systems, as it allows the identification of the most relevant terms and concepts in a given thematic domain. For these authors, domain analysis is a dynamic and continuous process, which must be constantly updated to keep up with changes and evolutions in the field of knowledge.
Naves (2000, p. 249)	The basic operation in subject indexing involves the indexer extracting the content from a document.	Krippendorff (2004).	Content analysis is a research technique for the objective, systematic, and quantitative description of the manifest content of communications.	Smiraglia (2011,2014)	It is characterized by the study of theoretical aspects of a given environment, generally represented by literature or a community of researchers, constituting a means for generating new knowledge about the interaction of a given scientific community with information.
Hjørland (2010)	Subject analysis can identify and describe the themes, concepts, and relationships present in a document or set of documents, aiming to enable their representation and subsequent retrieval.	Hsieh e Shannon (2005)	Content analysis is a research technique to uncover the themes, patterns, and underlying meanings that emerge from textual data.	López-Huertas (2015, p. 570)	It is a general proposal that has since served as a theoretical reference to address numerous studies in library and information science (LIS) in general and in knowledge organization (KO) in particular.
Gil (2010)	Subject analysis is a research technique that involves identifying and categorizing themes in a text, aiming to understand its structure and organization.	Morville (2005)	Content analysis is a research technique for identifying and analyzing the manifest and latent aspects of a text's content.	DIAS (2015, p. 7-8).	Domain Analysis is a process with the purpose of "identifying objects, processes and their relationships to build a vocabulary of a given domain that can be used for both organizational and information and knowledge retrieval activities".
Bardin (2011)	Subject analysis is a research technique that identifies the main themes and subthemes addressed in a text and categorizes them into predefined categories.	Braun e Clarke (2006)	Content analysis is a research technique aimed at systematically and objectively identifying, categorizing, and interpreting communication patterns within a specific textual corpus.		
Minayo (2014)	Subject analysis is a research technique that identifies and classifies the main themes addressed in a text based on predefined categories.	Altheide e Schneider (2013)	Content analysis is a research technique that aims to understand social practices by interpreting and analyzing texts, images, and other types of communicative data.		
Lima (2020)	Subject analysis is an intellectual, hence subjective, activity that begins with the technical reading of the main parts of a document. It requires linguistic, cognitive, and logical knowledge from the indexer to determine what a document is about, that is, its subject.				

Appendix B: Definitions of subject analysis, content analysis, and domain analysis proposed by authors in the field

Comparison table							
	Subject analysis		Content analysis		Domain analysis		
Definition	It is an intellectual activity, therefore subjective, which begins with the technical reading of the main parts of a document, requiring linguistic, cognitive, and logical knowledge from the indexer to be performed to determine what a document is about, that is, what its subject is. (LIMA, 2020, p.9)		It is a set of communication analysis techniques that uses systematic and objective procedures to describe the content of messages and indicators (quantitative or not) that allow the inference of knowledge related to the production/reception conditions (inferred variables) of these messages. (BARDIN, 1977).		It is a theoretical approach in Information Science (IS) that states that the best way to understand and information is to study knowledge areas as 'communities of discourse,' which are parts of the societal division of labour. (HJØRLAND; ALBRECHTSEN, 1995, p.23)		
Characteristics	It is a technique		It is a technique		It is a technique		
Objective	Represent the content of a document concisely to ensure retrieval.		Analyze data from communications, seeking to understand the meanings and senses of messages, going beyond a shared reading. In other words, it aims to identify what is being said about a particular topic and whether there is a need to decode what is being communicated.		Analyze the domain as a way to visualize the emergence and coherence of a domain and also as a means of mastering the parameters of the universe in which this domain operates.		
Functions	Representation of informational content to facilitate information retrieval.		Describe and interpret the data collected from the qualitative and quantitative analysis of occurrences of specific terms, constructions, and references in a given document within a theme.		Understanding and delimiting the information set within a specific context by identifying communication patterns and information relevance. Seeking insights to comprehend a domain or to conduct an epistemological approach or as support for the analysis of the addressed topic.		
Procedures	Technical reading and analysis	Technical reading and analysis	Pre-analysis	Skimming		10 procedures (Hjørland 2002)	Domain identification
				Selection of documents (corpus)			Identificação das principais teorias, conceitos e métodos
				Reformulation of hypotheses and objectives			Identification of relevant information sources
				Preparation of the material			Critical analysis of information sources
		Identification of concepts representing this content	Exploration of the material	Administration of techniques in the corpus	Critical analysis of key documents		
					Critical analysis of information sources		
			Selection of valid concepts for retrieval	Handling of results, inference, and interpretation (BARDIN, 1977).			Identification of key actors
							Critical analysis of key actors
Identification of gaps in the domain							
Synthesis of results							
Approach	Qualitative		Qualitative and Quantitative		Theoretical-methodological		
Source area	Librarianship		Linguistics		Computer Science/Information Science		
Main authors	FOSKETT(1973), LANCASTER (2004), MAI, WENSE ; NAVES (2007), FUJITA (2006,2007,2008,2009,2010); LIMA (2020, 2021) ABNT NBR 12676 (1992)		H. Lasswell (1927); Laurence Bardin (1977)		Hjørland e Albrechtsen (1995), Hjørland (2002, 2004...), Tennis (2003), Smiraglia (2012),		