applied sciences

MDPI

Article

Textual Analysis of the DigComp 3.0 Framework Using Frequency and TF–IDF Methods

Danial Zare *, Luis Fernández-Sanz, Vera Pospelova and Ines López-Baldominos

Department of Computer Sciences, University of Alcalá, 28801 Madrid, Spain; luis.fernandez.sanz@uah.es (L.F.-S.); vera.pospelova@uah.es (V.P.); ines.lopezb@uah.es (I.L.-B.)
* Correspondence: danial.zare@uah.es; Tel.: +34-657521640

Featured Application

This article identifies keywords and action verbs supporting manual interpretation of DigComp components by practitioners and non-experts, reveals inconsistencies in DigComp 3.0 textual formulations, and extends computational analysis research on digital competence frameworks.

Abstract

The recently released DigComp 3.0 framework represents the latest evolution of the European reference model for digital competences, expanding its descriptors, proficiency levels, and learning outcomes. This study applies text-mining and natural language processing (NLP) techniques to systematically analyse the textual content of the framework across all its dimensions, including competence descriptions, proficiency-level statements, and learning outcomes. After performing standard text pre-processing steps—such as normalization, tokenization, and filtering—we extracted general statistics on the structure of the framework and conducted keyword analysis using term frequency and TF–IDF (Term Frequency–Inverse Document Frequency) methods. These techniques enabled the identification of the most meaningful terms and action verbs that characterize the formulation of digital competences and their components in DigComp 3.0. These terms are considered a useful manual instrument for identification of DigComp elements in text according to the opinion of surveyed practitioners. The results also show inconsistency in the characterization of the DigComp components regarding the presence of these terms.

Keywords: skill frameworks; text mining; NLP; TF-IDF; DigComp 3.0

1. Introduction

Digital competence is the set of knowledge, skills, attitudes, abilities, strategies and awareness that is required when using ICT and digital media to perform tasks; solve problems; communicate; manage information; behave in an ethical and responsible way; collaborate; and create and share content and knowledge for work, leisure, participation, learning, socializing, empowerment and consumerism [1]. In this context, the Digital Competence Framework for Citizens (DigComp 3.0) [2], developed by the JRC, an agency of the European Commission, has emerged as a pivotal reference for defining, assessing, and enhancing digital literacy. DigComp outlines five core domains of digital competence, providing a comprehensive structure for individuals, educators, and policymakers to address the evolving demands of the digital age.

Check for updates

Received: 15 May 2026
Revised: 19 June 2026
Accepted: 20 June 2026
Published: 9 July 2026

Copyright: © 2026 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license.

The importance of the DigComp framework lies in its ability to standardize digital competences across diverse educational and professional contexts. One notable example is the "DigComp into Action" guide [3], which presents 38 inspiring practices of DigComp implementation across education, training, and employment sectors. Another example is the "DigComp at Work" report [4], which focuses on employability and employment contexts. By offering a unified taxonomy, DigComp enables the alignment of curricula, training programs, and policy initiatives with the skills required for active participation in modern society. Additionally, DigComp can serve as a foundation for addressing digital divides, promoting equity, and fostering lifelong learning, making it an indispensable tool for governments, institutions, and organizations aiming to prepare citizens for technological challenges.

Traditionally, the analysis and mapping of skills and competence frameworks have been conducted manually by researchers and domain experts. In such manual approaches, competence statements are individually reviewed, interpreted, and classified into relevant cognitive categories, skills clusters, or learning taxonomies. While these expert-driven processes offer valuable qualitative insights, they also come with significant limitations. Manual analysis is time-consuming; labour-intensive, involving expensive qualified experts; and highly dependent on subjective interpretation, which can lead to inconsistencies across studies. Moreover, as competence frameworks expand in size and incorporate complex relationships among items, while also evolving to address emerging technologies and labour market needs, manual methods become increasingly difficult to scale and update in a timely manner.

The advances in natural language processing (NLP) techniques and the rapid evolution of Large Language Models (LLMs) now enable automation of many aspects of framework analysis with greater speed, consistency, and scalability. These techniques not only accelerate the analysis process but also reduce the risk of subjective bias, making the analysis more transparent and replicable. Of course, it is essential that results from automated processes represent an input to experts and practitioners who can then check and/or customize to specific contexts [5]: they can analyse and compare skills frameworks more systematically, enabling data-driven insights into their structure, cognitive emphasis, and potential gaps that represent insights that are difficult to achieve through manual analysis alone.

This study contributes to examining the recently published DigComp framework [2], addressing a key dimension: easier advanced analysis of competences of the framework to provide insights for better identification of concepts and improvement in the consistency of terminology. This study also aims at demonstrating how text analysis and NLP methods can enhance our understanding of the digital competencies described by DigComp, support evidence-based revisions to competence frameworks, and provide actionable insights for policymakers, educators, and instructional designers. And finally, this article is the first study to apply a systematic text analysis to the DigComp framework, with the aim of examining its internal structure and terminology in a comprehensive and empirical manner.

Inspired by the problems associated with insufficient digital skills highlighted in recent studies on digitization in education (e.g., [6]), this article emphasizes the critical importance of digital skills and competences and a valid measure which is suitable across the life span. Digital skills are essential not only for effective participation in technology-enhanced learning environments, but also for ensuring equity, adaptability, and quality in educational outcomes.

The article is structured as follows. In the next section, we present a literature review focusing on the DigComp framework, synthesizing existing research and contextualizing its development, applications, and critiques. After that, we detail the proposed methodology

for the textual analysis of the DigComp framework. The Results section demonstrates the obtained results of the research. Subsequently, the Discussion section addresses the research questions and interprets the findings, considering their broader implications. Finally, Section 6 concludes the paper by summarizing key insights and exploring future research directions that could extend or refine this study.

2. Literature Review

The DigComp framework has undergone several revisions and updates since its initial release in 2013 [7]. The first version established a reference model to describe the digital competences required by European citizens [7]. In 2016, DigComp 2.0 was introduced, updating the conceptual definitions and terminology to align with rapid technological developments [8]. This was followed by DigComp 2.1, which introduced eight proficiency levels to better describe competence progression from foundational to highly advanced [9]. DigComp 2.2 expanded the framework further, offering new examples of knowledge, skills, and attitudes to reflect emerging technologies and societal changes [2]. Finally, in DigComp 3 [2], the framework was updated to better reflect emerging digital technologies and new societal challenges, including data, AI, and online safety. It also refined competence descriptors and proficiency levels to improve clarity, usability, and alignment with current educational and professional needs. More importantly, learning outcomes were added to the framework to clearly define what learners are expected to know, understand, and be able to do at each competence level.

This continuous evolution highlights DigComp's dynamic and adaptive nature, aiming at refining descriptions and concepts to better provide a comprehensive and practical framework for policymakers, educators, and citizens. Since its creation, DigComp has been widely adopted across Europe and beyond as a key policy and educational tool for enhancing digital competence among citizens. Various public institutions, including national governments, ministries of education, and labour agencies, have used DigComp as a benchmarking framework to design digital literacy strategies, training programs, and competence assessments [9]. For example, Spain incorporated DigComp into its National Plan for Digital Competences, aligning public training initiatives for lifelong learning and workforce development with DigComp's competence areas and proficiency levels [10]. Similarly, employment agencies in several EU countries use DigComp to define digital skills profiles needed for different jobs, helping to bridge the digital skills gap and guide job seekers in developing relevant competencies [3].

Beyond governmental and public-sector applications, DigComp has also been embraced by educational institutions, corporate training programs, and non-governmental organizations. In formal education, DigComp serves as a curriculum design tool, helping educators incorporate digital competence development into school programs, teacher training, and student assessment (e.g., see [3,11–14]).

Other EU frameworks derived from DigComp appeared to address needs in specific contexts. DigCompEdu, for example, focuses on digital competence for educators [15–17]. Similarly, DigCompOrg addresses digital competence within organizations and other related frameworks such as DigCompConsumers and EntreComp, which focus on digital competence for consumers and entrepreneurs, respectively [18]. These frameworks highlight the specific digital skills needed for navigating consumer markets and successfully engaging in entrepreneurial activities in a digital context. The literature also shows the interest in using DigComp-related frameworks as a basis for different digital competence assessment instruments [19], also combined with DigCompEdu [20].

Recent studies have increasingly explored the use of natural language processing (NLP) and text-mining techniques for the analysis and development of competency and skills

frameworks. Garman et al. [21] demonstrated the potential of NLP methods to support the development and examination of competency frameworks through the systematic analysis of competency statements and descriptors. More recently, Zare et al. [22] applied text-mining and NLP techniques to enrich IT professional skills frameworks, showing how automated textual analysis can reveal semantic relationships and support framework refinement. Furthermore, Zare et al. [23] employed transformer-based NLP models to perform semantic mapping between soft skills and standardized competency frameworks, illustrating the growing role of advanced language models in competence classification and alignment. These studies indicate a broader shift from manual examination of skills frameworks toward data-driven and automated approaches.

Frequency-based analysis and TF–IDF have also become widely used techniques for examining large collections of competence- and skills-related textual data. Recent studies have employed these methods to identify salient concepts, distinguish domain-specific terminology, and support the construction and refinement of skills taxonomies and competency repositories. For example, research on occupational and educational skill standards has demonstrated that TF–IDF can effectively highlight discriminative terms and reveal thematic differences across competence descriptions, learning outcomes, and skills classifications [24–26]. These approaches provide a transparent and interpretable means of analysing framework content and have proven useful for supporting framework comparison, enrichment, and automated classification tasks.

To the best of our knowledge, after conducting a thorough search on platforms such as Google Scholar, ScienceDirect, and other academic databases, we did not find any systematic content analysis, linguistic analysis, or computational text analysis that directly applies NLP and text-mining techniques to the DigComp 3.0 framework. However, Roh et al. [27] examined the integration of digital literacy in language arts curricula across Ontario, Australia, and South Korea, finding limited emphasis and weak connections among digital literacy components, with variations in focus across regions and improvements in revised standards, highlighting the need for more balanced and integrated inclusion of digital literacy skills. There is also an isolated case where text mining and web scraping were applied to curricula of vocational training centres in Germany to detect their consideration of digital skills in the programs, not working with DigComp [28]. Another study used manual text analysis assisted by a qualitative analysis tool to detect DigComp concepts within secondary school textbooks [29].

Other existing research overwhelmingly focuses on the framework's application in policy and education or examines the research production around DigComp through bibliometric analysis [30,31] or direct manual analysis [32] (e.g., such as manual coding or categorization interpretation [17]). While these studies provide valuable insights into the impact and uptake of DigComp, they do not explore what implicit patterns, terminological shifts, or conceptual structures exist within the framework's own textual evolution across its different versions. This gap in the literature highlights an important unexplored dimension of DigComp: understanding how the language, terminology, and semantic framing used within the framework itself have evolved and what new insights could emerge from analysing DigComp as a corpus using content analysis or NLP techniques.

Despite the extremely limited research applying basic text analysis techniques to the contents of the DigComp framework, similar methods such as NLP and text mining have already been applied to other competence frameworks to extract valuable insights from their content and textual structure. For instance, ref. [33] employs advanced NLP methods to enhance the ESCO classification by explicitly labelling soft skills, thereby improving its applicability in the modern labour market. Ref. [34] uses text mining to assess how well information systems curricula align with the SFIA framework, ensuring educational

programs meet industry standards. Ref. [35] discusses the use of LLMs to match job descriptions with ESCO skills, highlighting the potential of NLP in automating skills extraction. Ref. [24] presents SkillNER, an NLP-based tool designed to extract and map soft skills from various texts, facilitating the integration of soft skills into frameworks like ESCO. Such references demonstrate how NLP techniques can be used to analyse and enrich skills frameworks, revealing patterns in skill descriptions, identifying semantic overlaps between competencies, and even detecting evolving terminology over time. Ref. [33] integrates a structured literature review with advanced NLP methods to systematically identify and categorize soft skills within the ESCO framework, enhancing the clarity and comprehensiveness of soft skills classification. Ref. [34] employs text mining to assess how well academic curricula align with industry requirements by analysing the SFIA framework.

These examples suggest that computational text analysis can serve as a powerful tool not only for mapping the explicit content of a framework but also for uncovering implicit knowledge embedded within its textual content [36], unfortunately not directly delivered in machine-readable format [37]. Applying NLP methods to DigComp could provide new insights into its internal coherence, conceptual evolution, and alignment with emerging digital competence needs. This situation inspires this study to advance in the deep analysis of textual semantic content of DigComp while also showing the contribution of systematic and scientific methods for processing content of this framework, providing a more objective and data-driven approach to enhance and facilitate the understanding of its concepts. As a result, this study is aimed at two main research objectives:

RO1. Propose a refined NLP method for the effective analysis of the contents of the DigComp framework based on the use of frequent terms in each of its dimensions and components.

RO2. Determine conceptual insights based on the analysis of the most frequent terms in the textual contents of the DigComp framework.

3. Proposed Methodology

The adopted methodology relies first on a consideration of the type of additional information that can be useful for the understanding of concepts and components of DigComp. When standards and frameworks show the keywords and frequent terms associated with each concept in a consistent way, the studies confirm that users get better understanding of information [38] and also perceive more credibility [39]. As an additional confirmation, the authors launched a survey (https://bit.ly/DigComp-keywords accessed on 17 March 2025 till 8 April 2025) through the most popular online community for practice of DigComp (created by AllDigital) asking about the usefulness of several complements of information with keywords for a better understanding of the framework.

The survey collected background information about respondents, including their professional role, country, level of expertise regarding the DigComp framework, and familiarity with recent DigComp developments. In addition, participants were asked to evaluate the usefulness and representativeness of the extracted terms through four main questions measured on a five-point Likert scale. The survey was designed to gather expert-informed opinions and preliminary feedback on the outcomes of the textual analysis. Given the relatively limited number of respondents, the survey results should not be considered statistically representative of the broader digital skills community, nor were they used as primary evidence for the study's conclusions. Instead, they served as a complementary source of qualitative and quantitative insights that supported the interpretation of the frequency and TF–IDF findings. The main objective of this questionnaire was determining the perceived significance of textual analysis and key terms in the DigComp framework as

well as understanding the importance and meaning of identifying the most critical terms within each of its dimensions.

A total of 21 participants took part in this questionnaire, including 5 educators/practitioners, 5 researchers, 7 consultants/experts, and 4 IT professionals from 9 countries, namely Spain, Italy, Romania, Slovenia, the Netherlands, Latvia, Austria, Peru, and Palestine. Regarding their familiarity with the latest version and updates of the DigComp framework, 48% identified themselves as very familiar and up to date, 43% as somewhat familiar, and 10% (2 participants) as not familiar at all. Additionally, participants described their expertise in DigComp 2.2 as follows:

- A total of 38% had experience applying the DigComp framework in educational/training experiences.
- A total of 33% had extensively worked on adapting the DigComp framework to different contexts or advising others on its application.
- A total of 19% had conducted comprehensive research based on the framework and had developed new resources, tools, or methods.
- A total of 5% had fully read and studied the framework at least once.
- Regarding the most crucial question in the questionnaire:
- "Do you think that offering a list of the most frequent terms associated with one of the following DigComp elements would help less experienced readers better identify them in real practice?" The responses were as follows:
- Frequent terms and keywords for each competence (Dimension 2): 52.4% strongly agree and 42.9% agree.
- Frequent terms and keywords for each competence proficiency level (Dimension 3): 47.6% strongly agree and 47.6% agree, 4.8% disagree.
- Frequent verbs for each competence proficiency level (Dimension 3): 52.4% strongly agree and 33.3% agree.

Given that the opinion on providing these lists of key terms was almost unanimously favourable, this suggested us focusing on this aspect in this study.

3.1. Input Textual Data

Unfortunately, most of the standards and framework are only available in traditional documents, even protected against extraction of text, and not ready for being processed (e.g., structured databases or information files ready for easy text processing [37]). Although modern text crunching techniques made popular by LLMs [40] would suggest that this would not be a barrier, having structured clean information is an advantage for deep text analysis. In any case, considering that the textual data of the DigComp 3.0 framework are stored in a CSV file in tabular format, we extracted some interesting initial information, which we will present in the following section.

In DigComp 3.0, the framework is structured into 5 competence areas (Dimension 1), encompassing 21 competences (Dimension 2) for citizens: information and data literacy, communication and collaboration, digital content creation, safety, and problem solving. Each competence area includes a short description, which can vary from a simplified combination of its competence descriptions to more synthetic statements.

Unlike DigComp 2.2, DigComp 3.0 no longer includes separate examples for knowledge, skills, and attitudes. Instead, it introduces learning outcomes for each competence. Each learning outcome is annotated with three labels:

- Proficiency Level: basic, intermediate, advanced, or highly advanced.
- Dimension Type: indicating whether the outcome reflects knowledge, skill, or attitude.
- AI Label: specifying AI-implicit or AI-explicit.

For our analysis, we considered all available textual information for each competence, including descriptions, proficiency levels, and learning outcomes, as provided in the DigComp 3.0 Excel database. This structure enables a systematic extraction and analysis of both general and distinctive terms, capturing the conceptual and operational emphasis of the framework across all competence areas.

3.2. Basic Text Analysis Using the Database

The basic statistics on the amount of text in the descriptive part of the framework for each competence and level and also the number of learning outcomes across proficiency levels are shown in Table 1. Regarding the statements, each one has a lead verb that mainly determines the content.

Table 1. Number of statements at each proficiency level.

	Proficiency Levels					Learning Outcomes				
Competence	**Basic**	**Intermediate**	**Advanced**	**Highly Advanced**	**Total**	**Basic**	**Intermediate**	**Advanced**	**Highly Advanced**	**Total**
1.1	4	4	3	4	15	6	6	4	5	21
1.2	6	7	5	3	21	7	14	6	4	31
1.3	5	6	6	5	22	7	12	10	6	35
2.1	5	6	6	4	21	10	11	6	4	31
2.2	6	4	3	3	16	7	6	3	4	20
2.3	4	8	6	3	21	9	13	8	4	34
2.4	3	4	3	4	14	4	7	6	5	22
2.5	4	3	5	3	15	5	5	8	3	21
2.6	6	5	5	3	19	10	8	8	3	29
3.1	6	5	4	4	19	8	8	4	5	25
3.2	4	4	3	4	15	7	7	6	6	26
3.3	4	6	4	3	17	7	10	6	4	27
3.4	4	7	7	4	22	8	16	11	4	39
4.1	4	3	4	4	15	6	4	5	4	19
4.2	8	7	3	3	21	10	10	3	3	26
4.3	6	7	5	4	22	12	10	7	4	33
4.4	5	5	3	4	17	6	6	3	4	19
5.1	4	3	3	2	12	7	3	3	2	15
5.2	3	3	4	3	13	7	3	4	3	17
5.3	3	3	3	3	12	3	6	4	3	16
5.4	3	3	4	3	13	5	5	4	3	17
Total	97	103	89	73	362	151	170	119	83	523

Since each learning outcome in DigComp 3.0 is labelled using the concepts of skills, knowledge, and attitude, we calculated the number of labels corresponding to each of these concepts for each competence. See Table 2. For example, out of the 21 learning outcomes related to competence 1.1, 11 are of the skill type, 5 are of the knowledge type, and 5 are of the attitude type. In addition, the learning outcomes are also labelled using AI labels. AI labelling indicates whether AI is explicitly mentioned (AI-E) or not explicitly mentioned but implicitly relevant (AI-I) [2]. Table 2 also shows the number of learning outcomes by AI label type for each competence.

3.3. NLP Techniques

3.3.1. Text Pre-Processing

For text mining, textual data related to the DigComp framework were organized in an Excel file. A Python program (version 3.10.11) was employed to pre-process the textual data. The raw text data underwent several cleaning and transformation steps to ensure the

results were reliable and meaningful. A custom pre-processing function was developed to perform the tasks shown in Figure 1.

Table 2. Skill, knowledge or attitude type of learning outcomes as well as AI labels.

	S/K/A				AI Labels			
Competence	**Skill**	**Knowledge**	**Attitude**	**Total**	**AI-Explicit**	**AI-Implicit**	**AI Not Implicit or Explicit**	**Total**
1.1	11	5	5	21	1	20		21
1.2	9	18	4	31	10	20	1	31
1.3	17	12	6	35		18	17	35
2.1	12	15	4	31	4	20	7	31
2.2	9	7	4	20	1	14	5	20
2.3	11	16	7	34	2	23	9	34
2.4	11	5	6	22	7	13	2	22
2.5	8	9	4	21		11	10	21
2.6	10	15	4	29	2	22	5	29
3.1	12	7	6	25	5	16	4	25
3.2	11	9	6	26	3	20	3	26
3.3	6	16	5	27	4	19	4	27
3.4	10	23	6	39	21	9	9	39
4.1	7	8	4	19	2	9	8	19
4.2	8	14	4	26	3	11	12	26
4.3	12	14	7	33	1	21	11	33
4.4	6	8	5	19	1	16	2	19
5.1	7	3	5	15		3	12	15
5.2	9	4	4	17	1	14	2	17
5.3	8	5	3	16	1	15		16
5.4	5	4	8	17		16	1	17
Total	199	217	107	523	69	330	124	523

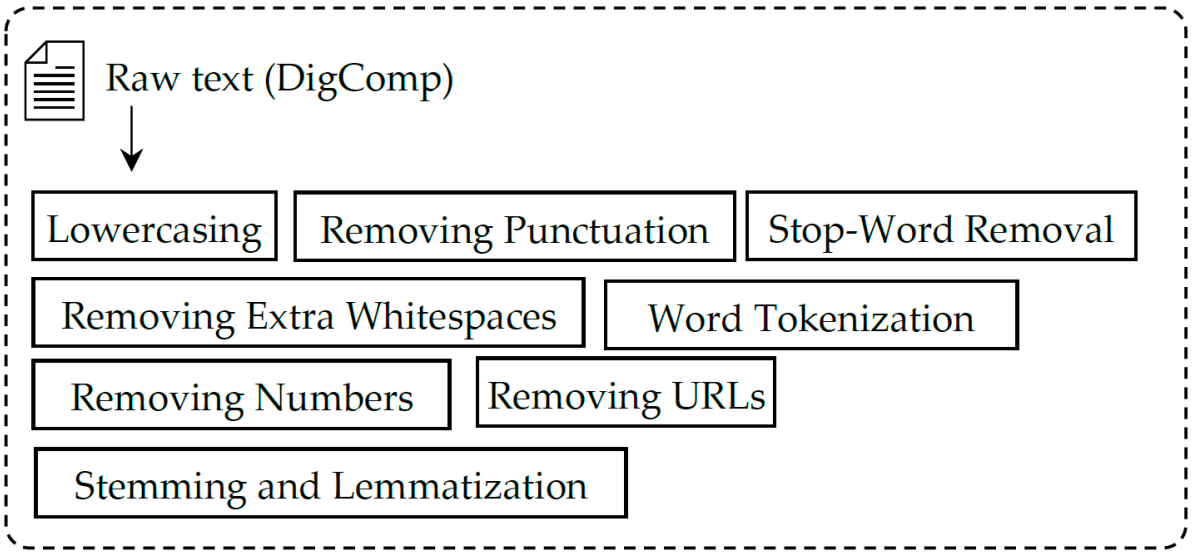


Figure 1. Pre-processing of the textual data.

It should be noted that stop-word removal was performed in Python using the Natural Language Toolkit (NLTK) library. Specifically, the standard English stop-word list provided by NLTK was applied during the pre-processing stage to filter out common functional words that do not contribute to the semantic meaning of the text. After tokenization and normalization (including lowercasing), each token was compared against the NLTK stop-word set, and non-informative words were removed from the corpus. In addition, a small number of domain-independent functional terms, consistent with the principles of

the NLTK stop-word collection, were manually excluded to further refine the textual data. This approach ensured a cleaner and more semantically focused dataset for subsequent TF–IDF analysis.

3.3.2. Part-of-Speech Tagging

Part-of-speech (POS) tagging is one of the most important tasks in the field of NLP. It assigns a POS tag to each word in a given sentence. For a short and simple sentence such as "*I like dogs*", a POS tagger can easily identify the word "*I*" as a pronoun, the word "*like*" as a verb, and the word "*dogs*" as a noun [41].

In this research, we use predefined functions from the NLTK library in Python along with POS tagging to extract selected words, such as the verbs used in the DigComp framework. Figure 2 illustrates the details of the workflow. Specifically, the textual data of each DigComp area are pre-processed separately, and after keyword extraction, POS tagging is applied to identify the grammatical categories of the words, such as nouns, verbs, and others.

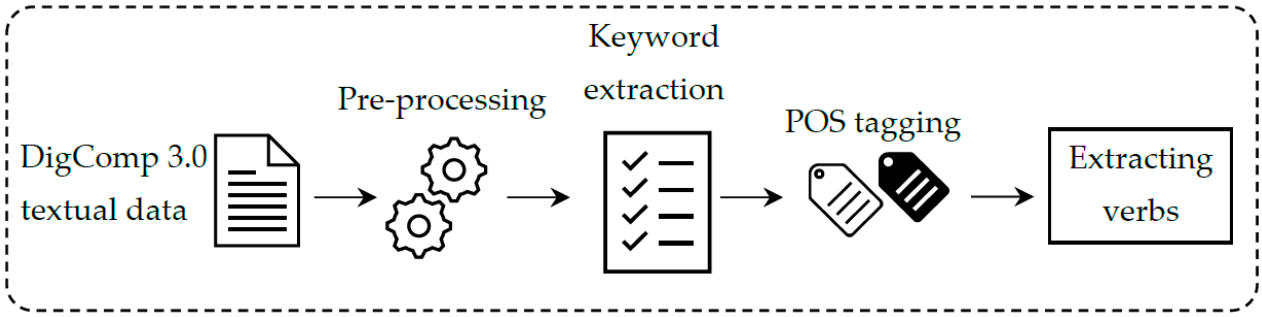


Figure 2. POS tagging.

3.3.3. TF-IDF

TF–IDF (Term Frequency–Inverse Document Frequency) is a widely used method in information retrieval and text mining to quantify the importance of a word within a document relative to a collection of documents (corpus) [42]. It combines two components:

- Term Frequency (TF): Measures how often a term appears in a specific document. Higher frequency indicates greater relevance within that document.
- Inverse Document Frequency (IDF): Measures how unique or informative a term is across the entire corpus. Terms that appear in many documents receive lower IDF values, while higher values are assigned to rare terms.

The TF–IDF score for a term is calculated as the product of its TF and IDF values. In this study, we applied TF–IDF to vectorize the textual data of the DigComp 3.0 framework using Python, enabling the identification of the most significant and distinctive terms across competences, proficiency levels, and learning outcomes.

For the TF–IDF computation, the textual corpus was first pre-processed using Python-based NLP libraries, specifically the Natural Language Toolkit (NLTK) and scikit-learn. These libraries were employed to perform standard text processing tasks, including tokenization, lowercasing, punctuation removal, and stop-word filtering. The processed documents were then transformed into a document–term matrix using scikit-learn's feature extraction module, from which TF–IDF scores were calculated for each term across the corpus. The implementation employed unigram tokens only (n-gram range = 1–1), allowing the analysis to focus on individual lexical units and facilitating the interpretability of the resulting keywords. Lowercase normalization was applied to avoid duplicate representations of the same term, while standard IDF smoothing (as implemented in scikit-learn's TF–IDF

vectorizer) was enabled to prevent division-by-zero issues and to reduce the influence of extremely rare terms. No sub-linear term-frequency scaling was applied, meaning that raw term frequencies were retained in the TF component of the calculation. The resulting TF–IDF values were subsequently used to rank terms according to their relative importance within each DigComp component. These parameter settings follow commonly adopted practices in exploratory text mining studies and support the transparency, reproducibility, and interpretability of the analysis.

The document units used for TF–IDF calculation depended on the specific scope of each analysis. When the objective was to analyse the terminology of the DigComp framework as a whole, all textual content extracted from the framework was considered as the analysis corpus. For component-specific analyses, separate corpora were created for each DigComp element, such as competence descriptions, proficiency levels, and learning outcomes. In these cases, the corresponding textual content was extracted and stored separately prior to TF–IDF computation. This approach enabled the identification of terms that were particularly representative of each framework component while maintaining consistency across analyses. For transparency and reproducibility, the exact input corpora used in each TF–IDF analysis are documented and made available in the accompanying Zenodo repository.

3.3.4. Analysis Process

Figure 3 illustrates the steps for obtaining the TF-IDF scores of words across different dimensions of the DigComp 3.0 framework. Using the Python program we developed, the textual data were first pre-processed (Figure 1), and then the TF-IDF scores of the extracted words were calculated using functions from the NLTK library and subsequently analysed.

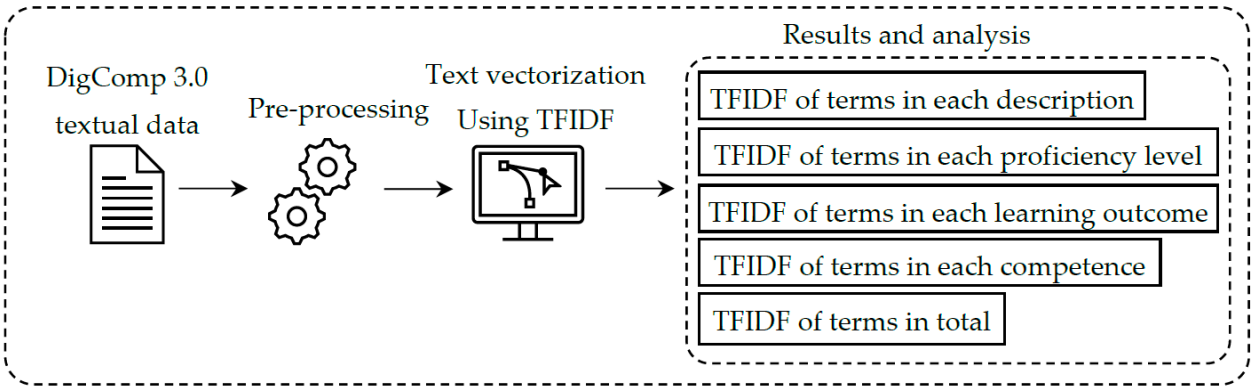


Figure 3. TFIDF analysis process.

It should be noted that the refinement proposed in this study does not refer to a novel TF–IDF formulation or a modified weighting scheme. Rather, it consists of the systematic integration of several established NLP techniques—including text pre-processing, part-of-speech tagging, term-frequency analysis, TF–IDF ranking, and cross-component comparison—within a unified analytical framework. This integrated approach enables the consistent examination of competence descriptions, proficiency levels, and learning outcomes, thereby providing a comprehensive and reproducible analysis of the textual structure of the DigComp 3.0 framework.

4. Results

This section presents the results of the process where the developed Python code fed the textual data from DigComp 3.0 into the program to generate the desired results. Here,

the results will be examined from three perspectives: analysis of verbs, term frequency and the TF-IDF score. The full table for each section is available at the Zenodo repository.

4.1. Analysis of Verbs

By analysing the verbs used in the descriptive part of the framework (specifically the area description and competence description), we extracted the verbs used in each area using NLP methods—such as first pre-processing the text and then applying POS tagging [43] with the spaCy library in Python. Table 3 presents the number of verbs used in each area. We examine the verbs used in the DigComp 3.0 framework from two perspectives: first, the number of verbs used in each area, and second, the number of verbs used at each proficiency level, disaggregated by each competence. In total, 213 unique verbs were extracted, some of which are presented in Table 3 by area.

Table 3. Number of verbs in each area.

Verb	Area 1	Area 2	Area 3	Area 4	Area 5	Sum
use	27	22	36	15	25	125
recognise	27	33	24	30	9	123
identify	15	33	20	24	25	117
manage	14	20	0	5	0	39
implement	7	7	1	9	0	24
refine	6	4	1	0	0	11
generate	6	4	8	2	0	20
influence	6	3	0	0	0	9
help	6	5	4	15	15	45
combine	6	4	2	0	0	12
update	5	0	0	0	5	10
⋮	⋮	⋮	⋮	⋮	⋮	⋮

As an example, the verb "*recognise*" has been used 27 times in area 1 while it has been used nine times in area 5, and subsequently it has been used 123 times across the whole framework. With a focus on verb usage across proficiency levels, Table 4 presents the distribution of verbs as they appear at each level throughout the DigComp framework.

Some of the verbs extracted from the DigComp 3 framework are highly generic auxiliary or functional verbs (e.g., *be*, *have*, *can*, *may*, *need*, *make*, *take*, *give*, *seem*, *exist*, *depend*). These verbs mainly serve grammatical or modal functions and do not convey domain-specific actions or competencies. Including them in an analytical table of verbs may add frequency noise without contributing meaningful insight into the conceptual structure of the framework. For the purpose of content-focused or competency-oriented text analysis, it is therefore reasonable to exclude such general verbs and focus instead on action-oriented, semantically rich verbs (e.g., *identify*, *evaluate*, *manage*, *implement*, *collaborate*), which more accurately reflect the skills, processes, and learning outcomes emphasized in DigComp.

For instance, verb "*recognise*" has been used 34 times at the basic level throughout the DigComp framework, while it has not been used at the highly advanced level at all.

Although some verbs (typically the less frequent ones) occur only in a single area, it seems that verbs alone are not sufficient to determine the competence area to which a statement belongs. This is logical as verbs are more related to proficiency levels than to areas because levels, as declared by the DigComp document, are inspired by Bloom's taxonomy [44]. Therefore, a more sophisticated NLP-based analysis is required to identify potential key terms that more accurately link a descriptive text to a competence area.

Table 4. Distribution of verbs across proficiency levels.

Verb	Basic	Intermediate	Advanced	Highly Advanced
acknowledge	20	12	2	1
identify	22	14	4	2
drive	1	0	0	0
recognise	34	17	1	0
implement	5	3	2	2
refine	1	2	1	1
update	3	1	0	0
support	8	7	14	14
help	8	3	5	5
create	2	4	1	0
develop	1	4	11	8
lead	0	0	0	18
contribute	0	1	2	23
promote	0	0	1	8
⋮	⋮	⋮	⋮	⋮

4.2. Frequency Analysis of Terms

When analysing the word frequency in the DigComp framework, a number of terms can be considered uninformative or low value for semantic text analysis and therefore may be excluded. These include function words, auxiliaries, and highly generic terms such as *one*, *other*, *way*, *well*, *main*, *possible*, *able*, *due*, *may*, and *exist*, which primarily serve grammatical or structural roles rather than conveying conceptual meaning. In addition, framework-defining or context-obvious terms such as *digital*, *technology*, *content*, *environment*, and *AI* appear with very high frequency but offer limited analytical discrimination because they are intrinsic to the DigComp domain itself. Retaining such terms can obscure more informative patterns, while filtering them out allows greater focus on conceptually meaningful words related to competencies, processes, ethics, safety, collaboration, and problem solving, which better capture the substantive intent of the framework. However, we have not excluded these terms in the presented tables.

Following the steps described in Figure 1, term frequencies are presented from several perspectives:

- The term frequency within the description of each competence;
- The term frequency at each proficiency level within every competence;
- The term frequency in the learning outcomes of each competence;
- The overall term frequency for each competence;
- The overall term frequency across the entire DigComp framework.

Table 5 shows the term frequencies used in the descriptions of competences 1.1, 2.3, and 4.2. "Fr" in the table refers to the frequency of each term. It is important to note that, for the purpose of calculating these term frequencies, the competence area name, competence area descriptor, competence name, and description associated with that competence were all processed and included in the analysis.

Table 5. Term frequency across descriptions.

Competence 1.1		Competence 2.3		Competence 4.2	
Term	**Fr**	**Term**	**Fr**	**Term**	**Fr**
inform	8	digital	9	digital	6
digital	7	technology	4	privacy	5
search	5	environment	3	person	5
content	5	one	3	data	5
environment	3	participate	3	use	4
source	2	communicate	2	environment	4
relevant	2	choose	2	technology	3
need	2	collaborate	2	aware	3
⋮		⋮		⋮	

Using a similar approach, we examined the term frequencies at the proficiency-level layer. For example, Table 6 presents the term frequencies for competence 1.1, broken down by the basic, intermediate, advanced, and highly advanced proficiency levels.

Table 6. Frequency of terms at proficiency levels for competence 1.1.

Competence 1.1							
Basic	**Fr**	**Intermediate**	**Fr**	**Advanced**	**Fr**	**Highly Advanced**	**Fr**
search	7	digital	4	digital	3	digital	4
use	4	search	4	search	3	search	4
digital	4	tool	2	tool	2	complex	3
tool	4	strategy	2	capable	2	specialize	3
inform	2	appropriate	2	continue	1	environment	2
depend	2	inform	2	strategy	1	inform	2
acknowledge	1	need	2	develop	1	need	1
⋮		⋮		⋮		⋮	

Regarding the learning outcomes (the newly added component of the DigComp framework), Table 7 presents the term frequencies at the learning-outcomes layer for competences 1.1, 2.3, and 4.2. "LO" in the following table stands for learning outcome.

Table 7. Frequency of terms in learning outcomes.

LO 1.1		LO 2.3		LO 4.2	
Term	**Fr**	**Term**	**Fr**	**Term**	**Fr**
search	24	digital	28	data	21
digital	19	participate	13	privacy	15
tool	10	technology	12	person	13
inform	7	platform	8	digital	10
result	6	civic	7	recognize	7
output	6	recognize	7	protect	7
use	5	right	5	environment	5
strategy	4	describe	5	legislation	5
complex	4	service	5	share	5
need	4	social	5	relate	4
specialize	3	individual	4	online	4
⋮		⋮		⋮	

This time, we examined the term frequencies within each competence individually. In other words, the textual data for each competence—including the area name, competence title, competence description, proficiency levels, and learning outcomes—was pre-processed and analysed as a unified set. Table 8 presents the term frequencies for competences 1.1, 2.3, and 4.2.

Table 8. Term frequency across each competence.

Competence 1.1		Competence 2.3		Competence 4.2	
Term	**Fr**	**Term**	**Fr**	**Term**	**Fr**
search	46	digital	57	data	44
digital	41	participate	27	privacy	33
inform	21	technology	25	person	31
tool	20	platform	15	digital	26
need	10	civic	15	protect	17
use	10	recognize	11	environment	15
result	9	right	10	recognize	14
strategy	8	service	9	share	11
output	8	use	9	use	10
complex	8	describe	9	legislation	10
⋮		⋮		⋮	

Finally, Table 9 presents the term frequencies across the entire text of the DigComp 3.0 framework.

Table 9. Term frequency across the entire text of DigComp 3.0.

Term	Fr	Term	Fr	Term	Fr	Term	Fr
digital	3640	create	521	share	247	skill	204
technology	1324	well-being	490	develop	241	right	193
content	1194	one	478	social	231	participate	192
environment	1173	collaborate	474	process	231	safety	192
ai	1010	communicate	468	interact	231	act	186
use	974	data	448	privacy	223	program	183
inform	964	respond	444	solve	219	recognize	183
implicit	851	problem	416	device	208	include	179
identify	712	support	406	technique	205	⋮	⋮
manage	555	aware	388	system	205		

4.3. TFIDF Analysis

As before, we examined the TF-IDF values of the terms in the DigComp 3.0 framework from previous perspectives (a–d in Section 4.2). Table 10 presents the TF-IDF scores of the terms exclusively within the description layer. These terms can be considered the most important words used in the description of each competence. The terms are sorted based on their TF-IDF values (indicating their level of importance).

Using a similar approach, we calculated the TF-IDF scores of the terms within the proficiency-level layer for each competence. Table 11 presents the most important terms used across the four proficiency levels—basic, intermediate, advanced, and highly advanced—for competence 1.1.

Table 12 also presents the TF-IDF scores of the terms within the learning-outcomes layer, which has been newly added to the framework. It is important to note that the TF-IDF values were calculated separately for each competence, considering only the textual data available in this specific layer.

Table 10. TFIDF scores of terms in competences 1.1 and 2.3.

Competence 1.1		**Competence 2.3**	
Term	**TFIDF**	**Term**	**TFIDF**
inform	0.5381	digital	0.6727
digital	0.4709	technology	0.2990
search	0.3363	one	0.2242
content	0.3363	participate	0.2242
environment	0.2018	environment	0.2242
need	0.1345	aware	0.1495
relevant	0.1345	choose	0.1495
evaluate	0.1345	collaborate	0.1495
manage	0.1345	communicate	0.1495
source	0.1345	society	0.1495
articulate	0.1345	assert	0.1495
⋮		⋮	

Table 11. TFIDF scores of terms at the proficiency levels for competence 1.1.

Competence 1.1							
Basic	**TFIDF**	**Intermediate**	**TFIDF**	**Advanced**	**TFIDF**	**Highly Advanced**	**TFIDF**
search	1.183	search	0.854	search	0.735	digital	0.857
digital	0.806	digital	0.854	digital	0.735	search	0.857
use	0.71	tool	0.712	capable	0.636	specialize	0.735
tool	0.71	strategy	0.591	tool	0.533	complex	0.735
inform	0.51	inform	0.582	other	0.475	inform	0.649
depend	0.486	appropriate	0.582	develop	0.475	environment	0.645
implement	0.387	need	0.582	assist	0.475	stay	0.49
basic	0.387	result	0.578	address	0.344	develop	0.49
refine	0.387	purpose	0.457	inform	0.344	technology	0.49
update	0.387	new	0.457	strategy	0.344	new	0.409
⋮		⋮		⋮		⋮	

Table 12. TFIDF scores of terms in the learning-outcome layer for competences 1.1 and 2.3.

Competence 1.1		**Competence 2.3**	
Term	**TFIDF**	**Term**	**TFIDF**
search	3.3814	digital	3.9486
digital	2.7864	participate	3.0528
tool	2.2143	technology	2.5284
inform	1.9831	civic	2.1012
output	1.7672	platform	2.0518
result	1.7672	recognize	1.9070
use	1.4663	describe	1.5950
strategy	1.3057	right	1.5863
need	1.2822	service	1.5290
complex	1.2492	social	1.4849
relevant	1.1880	environment	1.4565
⋮		⋮	

By considering all textual data available for each competence in the framework (including the competence title, description, proficiency levels, learning outcomes, etc.), Table 13 presents the TF-IDF scores of the terms for competences 1.1 and 2.3.

Table 13. TFIDF scores of terms within competences 1.1 and 2.3.

Competence 1.1		**Competence 2.3**	
Term	**TFIDF**	**Term**	**TFIDF**
content	0.5045	one	0.3168
source	0.3743	digital	0.2907
inform	0.2527	environment	0.2890
digital	0.2071	limit	0.2113
store	0.1872	share	0.2113
retrieve	0.1872	collaborate	0.2113
process	0.1872	manage	0.2113
organize	0.1872	reputation	0.2113
manage	0.1872	general	0.2113
locate	0.1872	feature	0.2113
judge	0.1872	diver	0.2113
general	0.1872	culture	0.2113
data	0.1872	present	0.2113
⋮		⋮	

Finally, by considering all textual data in the DigComp 3.0 framework, we calculated the TF-IDF scores of the terms, as shown in Table 14. These terms can therefore be regarded as the most significant words used throughout the entire framework.

Table 14. TFIDF scores of terms in DigComp 3.0 (in total).

Term	**TFIDF**	**Term**	**TFIDF**	**Term**	**TFIDF**	**Term**	**TFIDF**
digital	0.683	create	0.098	aware	0.073	protect	0.053
technology	0.249	well	0.098	advance	0.072	intermediate	0.052
content	0.224	being	0.092	knowledge	0.065	compute	0.052
environment	0.220	one	0.09	evaluate	0.063	ethic	0.051
ai	0.190	collaborate	0.089	other	0.063	improve	0.049
use	0.183	communicate	0.088	need	0.061	impact	0.049
inform	0.181	data	0.084	search	0.061	person	0.047
implicit	0.160	respond	0.083	explicit	0.058	integrate	0.047
identify	0.134	problem	0.078	apply	0.057	general	0.047
manage	0.104	support	0.076	basic	0.056	⋮	⋮

5. Discussion

In this section, we will take a closer look at the results obtained in the previous section. As previously mentioned, due to space limitations, it is not possible to present all the results and provide a comprehensive analysis of the DigComp framework across all dimensions and layers within this paper. Nevertheless, we will focus on analysing only the results that were presented earlier.

The findings presented in this study should be interpreted primarily as observations derived from quantitative textual analysis rather than as definitive evidence of conceptual quality or effectiveness of the DigComp 3.0 framework. The identified frequency patterns, TF–IDF rankings, and variations in terminology usage provide indications of how competences, proficiency levels, and learning outcomes are linguistically structured and

differentiated within the framework. While these patterns may suggest varying degrees of terminological consistency and practical interpretability, such conclusions should be considered indicative rather than conclusive. Further validation through expert evaluation, comparative framework analysis, and empirical applications in educational and professional contexts would be valuable to confirm the practical implications of the observed textual characteristics.

5.1. Discussion on the Use of Verbs in the Proficiency Levels

The distribution of verbs across proficiency levels in Table 4 provides partial empirical support for the claim that DigComp 3.0 is inspired by Bloom's taxonomy. Verbs typically associated with lower cognitive processes in Bloom, such as *recognise*, *identify*, and *acknowledge*, are predominantly concentrated at the basic level, while verbs linked to higher-order cognition, such as *assess*, *evaluate*, *diagnose*, *design*, *develop*, and *lead*, appear mainly at the advanced and highly advanced levels. This general progression suggests that DigComp adopts Bloom's underlying principle of increasing complexity and sophistication of performance. However, this alignment remains broad and indicative rather than systematic, pointing to inspiration rather than direct operationalization of Bloom's hierarchical model.

A major contradiction revealed by the table is that the same verbs recur across multiple proficiency levels, undermining the assumption that verb choice alone signals cognitive or skill progression. Verbs such as *use*, *support*, *help*, *manage*, and *share* appear consistently from basic through highly advanced levels, often with substantial frequencies at higher levels. This suggests that in DigComp, proficiency is not determined by the verb itself but by the complexity of the task, the context of use, and the degree of autonomy or responsibility involved. Consequently, DigComp diverges from Bloom's verb-centred logic, where specific verbs are tightly coupled with specific cognitive levels.

A second contradiction emerges from the presence of verbs traditionally associated with higher levels of Bloom's taxonomy—such as *create*, *generate*, and *develop*—within the basic and intermediate levels. In Bloom's framework, these verbs are reserved for advanced cognitive processes, yet in DigComp they may describe relatively simple or guided forms of creation. This indicates that DigComp differentiates proficiency not by the presence of creative activity per se, but by its scope, impact, and sophistication. As a result, Bloom's cognitive hierarchy cannot fully account for how DigComp conceptualizes progression in digital competence.

The third contradiction concerns the overall logic of progression in DigComp, which appears to be driven more by role, responsibility, and societal impact than by purely cognitive development. The strong concentration of verbs such as *lead*, *contribute*, *promote*, and *support* at the highly advanced level suggests a shift from individual task performance toward influence, coordination, and leadership within digital environments. This trajectory does not map neatly onto Bloom's cognitive sequence but instead reflects a socio-technical and competence-based model of growth. Taken together, these findings suggest that while DigComp 3.0 draws linguistic and conceptual inspiration from Bloom's taxonomy, its verb usage lacks full internal consistency. Therefore, it would be recommendable that a future revised version of DigComp explicitly, if deemed possible, refines and recalibrates the verbs and performance statements associated with each proficiency level to improve conceptual clarity, coherence, and alignment between language and intended competence progression.

5.2. Discussion on the Frequency of Terms

The frequency analysis reveals several patterns in the textual structure of DigComp 3.0 across different layers of the framework. As shown in Table 5, competence descriptions consistently emphasize core concepts such as *digital*, *inform*, *search*, and *environment*. These

high-frequency terms indicate that even at the description level, DigComp relies on a stable vocabulary centred on information handling, digital contexts, and user interaction. The recurrence of these terms across competences 1.1, 2.3, and 4.2 highlights the shared conceptual foundations underlying different competence areas. Therefore, it is relevant to note that the presence of these terms in other documents (e.g., training syllabi) will not be helpful for differentiating areas with a simple manual analysis.

A similar pattern emerges at the proficiency-level layer. In Table 6, highly generic and transversal terms such as *search*, *digital*, *tool*, and *inform* dominate across all proficiency levels for competence 1.1. These terms are inherently foundational and therefore not easily grouped by proficiency level. Rather than introducing distinct lexical sets, proficiency progression mainly refines the same conceptual vocabulary through qualifiers such as complex, specialize, or improve. While this reflects an increase in cognitive and operational complexity, it also limits the usefulness of DigComp for practitioners who attempt to manually infer proficiency levels from external documents (e.g., syllabi), as the same verbs and terms recur across levels. From this perspective, future revisions of DigComp could consider complementing transversal vocabulary with more level-sensitive linguistic markers to facilitate practical interpretation, in line with the spirit—though not the strict structure—of taxonomies such as Bloom's [44].

The learning-outcomes layer shows a broader distribution of action-oriented terms. In Table 7, items such as *search*, *digital*, *data*, *privacy*, *and participate* emerge as the most frequent. Compared to descriptions and proficiency statements, learning outcomes include more behavioural and context-driven vocabulary (e.g., *platform*, *legislation*, *service*, *empower*), reflecting the practical orientation of these statements and the need to articulate observable learner performance. However, the learning outcomes are far from offering clearly segmented terminology patterns or being decisive in identifying them in other contexts or documents.

When analysing the entire textual content of each competence (Table 8), the internal consistency of the terminology becomes more evident. For instance, competence 1.1 is strongly associated with *search*, *digital*, and *inform*, whereas competence 2.3 centres around *digital*, *participate*, and *technology*, and competence 4.2 prominently features *data*, *privacy*, and *person*. This layer-integrated view demonstrates that each competence maintains a distinctive lexical profile, even while sharing high-frequency foundational terms with the broader framework. This pattern could be helpful for its practical application in mapping of other documents to DigComp competences.

Finally, the corpus-level analysis in Table 9 underscores the overwhelming prominence of the term *digital*, which appears 3640 times—far exceeding any other item. Other frequently occurring terms such as *technology*, *content*, *environment*, *use*, and *inform* reflect the conceptual pillars of the framework. The presence of terms like *AI*, *well-being*, *collaborate*, and *protect* in the top frequency range also indicates DigComp 3.0's broadened scope, which now incorporates emerging technologies, human-centred dimensions, and responsible digital practices. Obviously those very common terms are useless when trying to identify items and components in documents or when mapping other documents to DigComp.

Overall, the frequency results across Tables 5–10 show strong lexical coherence between layers, a consistent reliance on a set of core terms, and clear differentiation between competences through domain-specific vocabulary. These patterns suggest that DigComp 3.0 is both conceptually integrated and semantically structured to support progressive competence development, but it is far from being useful for basic identification and segmentation of components in other texts only looking at frequent terms. So, aid to more basic, even manual, identification of DigComp components in other documents is not well

supported and more sophisticated analysis with NLP would be needed, something not available for most of the practitioners.

5.3. Discussion on the TFIDF Score

The TF–IDF analysis provides further insight into how the DigComp 3.0 framework distinguishes the semantic importance of terms across its layers. Unlike raw frequency, which captures repetition, TF–IDF highlights the words that are comparatively unique or conceptually defining within each competence or layer.

At the description level, Table 10 shows that competences develop distinct lexical profiles based on their specific focus areas. For example, *inform*, *digital*, and *search* have the highest TF–IDF values for competence 1.1, indicating their centrality to this competence's meaning. In contrast, competence 2.3 emphasizes *digital*, *technology*, and *participate*, reflecting its civic-orientation and participation-oriented nature. TF–IDF therefore exposes conceptual differentiation that simple frequency counts could not fully capture.

A clearer stratification appears within the proficiency levels. As shown in Table 11, the most salient terms shift as proficiency increases. For competence 1.1, *search* and *digital* maintain high importance across all levels, but advanced stages highlight terms such as *capable*, *complex*, and *specialize*. These higher-level terms reveal the framework's progression from basic operational skills toward more strategic, analytical, and specialized behaviours. TF–IDF thus reflects how lexical signalling aligns with cognitive and performance complexity.

In the newly introduced learning-outcomes layer, Table 12 demonstrates a much more pronounced weighting of key terms. For competence 1.1, terms such as *search*, *digital*, *tool*, and *inform* have notably high TF–IDF values, suggesting that learning outcomes depend heavily on a concentrated set of action-oriented and measurable terms. Similarly, competence 2.3 highlights *digital*, *participate*, and *technology*, underscoring the behavioural emphasis of this layer. Compared with descriptions and proficiency levels, learning outcomes show sharper term importance peaks, consistent with their role in specifying observable learner performance.

When aggregating all textual data for each competence (description + proficiency levels + learning outcomes), Table 13 reveals a more balanced but still distinctive set of TF–IDF-weighted terms. For competence 1.1, the leading terms—such as *content*, *source*, and *inform*—reflect a strong orientation toward information processing and evaluation, while competence 2.3 highlights terms such as *one*, *digital*, and *environment*, signalling its broad socio-digital scope.

At the corpus level, Table 13 identifies the most important terms in the entire DigComp 3.0 framework. The term *digital* unsurprisingly achieves the highest TF–IDF score, consistent with its central role in the framework's conceptualization of digital competence. Other highly weighted terms such as *technology*, *content*, *environment*, and *ai* represent the broader thematic pillars of DigComp 3.0, including technological interaction, contextual awareness, and emerging digital trends. Notably, TF–IDF values for behavioural or transversal terms (e.g., *collaborate*, *communicate*, *support*) are comparatively lower, suggesting that while such terms are frequent, they are more evenly distributed across competences and thus carry lower discriminative power.

Overall, the TF–IDF analysis across Tables 9–13 demonstrates that DigComp 3.0 possesses a well-structured lexical architecture: competences show distinct conceptual signatures, proficiency levels emphasize developmental progression, learning outcomes concentrate meaning through high-salience operational terms, and the full framework reflects a core vocabulary aligned with its digital competence objectives. However, as commented on above, the segmentation of terms for the identification of components is not possible with basic methods.

The proposed NLP-based approach complements traditional expert analysis by providing a reproducible, scalable, and quantitatively grounded mechanism for examining competence frameworks. While expert review remains essential for contextual interpretation, automated text analysis enables the systematic identification of salient terms, linguistic patterns, and potential inconsistencies across framework components. Such analyses can be repeated under identical conditions and applied to large collections of competence descriptions with minimal additional effort.

5.4. Analysis of Research Objectives

Research Objective 1 (RO1) sought to develop a refined NLP method capable of effectively analysing the textual contents of the DigComp 3.0 framework using frequent and distinctive terms across its dimensions and components. The multi-layered analyses—conducted separately for descriptions, proficiency levels, learning outcomes, individual competences, and the entire framework—demonstrate that combining frequency counts with TF–IDF weighting provides a comprehensive and discriminative representation of DigComp's conceptual structure. The results from Tables 4–8 (frequency) and Tables 9–13 (TF–IDF) show that each layer contributes unique lexical signals: descriptions rely on a stable foundational vocabulary (e.g., *digital*, *inform*, *search*), proficiency levels reflect lexical progression aligned with increasing complexity, and learning outcomes emphasize highly action-oriented and outcome-specific terms. The refined NLP pipeline—consisting of pre-processing, normalization, multi-layer frequency analysis, and TF–IDF scoring—thus proves effective in capturing both the shared terminology across the framework and the distinctive terms that characterize each competence. This demonstrates the method's suitability for structured, scalable, and reproducible content analysis of digital competence frameworks.

Research Objective 2 (RO2) aimed to identify conceptual insights revealed through the analysis of the most frequent and discriminative terms used in DigComp 3.0. The results highlight several important conceptual patterns. First, across all layers, the framework consistently centres on terms such as *digital*, *technology*, *environment*, *inform*, and *use*, indicating a strong emphasis on digital information management, technological interaction, and contextual awareness. Second, TF–IDF analysis reveals competence-specific conceptual signatures—for example, competence 1.1 is closely associated with *search*, *inform*, and *content*, while competence 2.3 is characterized by terms such as *digital*, *participate*, and *technology*. Third, learning outcomes show sharper peaks in term importance, underscoring the framework's focus on measurable, action-oriented behaviours. Finally, corpus-level TF–IDF results indicate that emerging concepts such as *AI*, *well-being*, and *privacy* have increasing conceptual weight in DigComp 3.0. Taken together, these findings suggest that DigComp exhibits a mixed capacity for conceptual segmentation: while high-frequency transversal terms promote coherence across the framework, competence-level and learning-outcome analyses provide sufficient lexical differentiation to support meaningful segmentation where it is most analytically and pedagogically relevant. Together, these insights demonstrate that term-based analysis provides a clear view of the framework's thematic structure, developmental logic, and evolving priorities—thus fulfilling RO2 by offering a coherent and empirically grounded interpretation of DigComp's conceptual landscape.

6. Conclusions and Future Works

This study demonstrates the effectiveness of text-mining approaches—specifically frequency analysis and TF–IDF—in uncovering the structural and conceptual patterns embedded in the DigComp 3.0 framework. By analysing the textual content across descriptions, proficiency levels, learning outcomes, individual competences, and the corpus as a

whole, the findings highlight a consistent lexical architecture shaped around core terms such as digital, technology, environment, inform, and search. These methods not only reveal the thematic coherence of the framework but also illuminate how DigComp distributes conceptual emphasis across its components. In particular, the analyses underscore that learning outcomes concentrate highly salient action-oriented terminology, while proficiency levels signal progression from basic operational tasks to increasingly strategic and analytical behaviours. Such insights confirm the value of refined NLP techniques for systematically characterising the internal logic and semantic structure of digital competence frameworks.

A notable contribution of this research is the identification of distinctive vocabulary patterns associated with different competence layers and proficiency levels. These patterns provide evidence of the hierarchical organization of DigComp, suggesting clear developmental trajectories from information acquisition to evaluation, problem solving, and responsible digital engagement. The results offer a data-driven basis for clarifying learning outcomes, strengthening vertical alignment across proficiency levels, and informing the design of more precise digital competence learning objectives and assessment tools.

Looking ahead, several promising research directions emerge. Analysis of existing training syllabi with the final engagement of experts and practitioners for interpreting results in identification of areas, competences, levels and learning outcomes would be extremely practical for the mapping of existing training offers to DigComp terminology. The extracted keywords in this study could be used for quick mapping to curricula with other sophisticated methods, as declared in [27]. This would enhance the practical relevance of the extracted insights and will support evidence-based refinements of the framework. As digital transformation continues to reshape education, work, and society, integrating data-driven analyses with expert perspectives will remain essential for ensuring that future iterations of DigComp effectively support learners in developing the competencies needed for a rapidly evolving digital landscape.

Furthermore, expanding the analysis to additional versions of the DigComp family—such as DigComp 2.2, DigCompEdu, and DigCompOrg—would help trace the evolution of digital competence conceptualization across educational and organizational contexts. Incorporating complementary NLP techniques, including topic modelling, clustering, and semantic similarity analysis, could enrich the interpretation of conceptual relationships within and across competences.

Finally, a larger manually annotated benchmark would provide an additional level of validation, and the authors consider this a valuable direction for future research.

Author Contributions: All authors contributed to the study conception and design as follows: D.Z.: Writing—original draft, Visualization, Formal analysis, Data curation, Investigation. L.F.-S.: Project administration, Writing—review and editing, Conceptualization, Supervision. V.P.: Writing—review and editing, Validation, Formal analysis. I.L.-B.: Writing—review and editing, Validation. All authors have read and agreed to the published version of the manuscript.

Funding: This research received no external funding.

Institutional Review Board Statement: According to the guidelines of the Comité de Ética de Investigación y Experimentación Animal of UAH (CEI-EA) (Committee of Ethics in Research and Animal Experimentation), which works under the principles of Declaration of Helsinki, studies based on fully anonymous surveys (i.e., non-anonymized data were not used), in which no personal data can be identified, do not require formal ethical approval or authorization from the committee. Prior to participation, all respondents were provided with clear information regarding the purpose of the study, and their voluntary participation was confirmed. Informed consent was obtained from all participants before they took part in the study.

Informed Consent Statement: Informed consent was obtained from all subjects involved in the study.

Data Availability Statement: The results of this study are available as Excel files in the Zenodo repository at https://zenodo.org/records/18311821 (accessed on 23 January 2026).

Acknowledgments: The authors would like to express sincere appreciation to all participants who took part in the survey. Their contributions and insights were essential to the completion of this research.

Conflicts of Interest: The authors declare no conflicts of interest.

Abbreviations

The following abbreviations are used in this manuscript:

NLP	Natural Language Processing
TF-IDF	Term Frequency–Inverse Document Frequency
DigComp	Digital Competence Framework for Citizens
LLM	Linear Large Language Model

References

1. Ferrari, A.; Punie, Y.; Redecker, C. Understanding Digital Competence in the 21st Century: An Analysis of Current Frameworks. In *21st Century Learning for 21st Century Skills*; Ravenscroft, A., Lindstaedt, S., Kloos, C.D., Hernández-Leo, D., Eds.; Lecture Notes in Computer Science; Springer: Berlin/Heidelberg, Germany, 2012; Volume 7563, pp. 79–92.
2. Cosgrove, J.; Cachia, R. *DigComp 3.0 European Digital Competence Framework*; Publications Office of the European Union: Luxembourg, 2025. Available online: https://data.europa.eu/doi/10.2760/0001149 (accessed on 17 December 2025).
3. Kluzer, S.; Pujol Priego, L. *DigComp into Action, Get Inspired Make It Happen: A User Guide to the European Digital Competence Framework*; Publications Office of the European Union: Luxembourg, 2018. Available online: https://data.europa.eu/doi/10.2760/112945 (accessed on 21 February 2025).
4. Centeno, C. *DigComp at Work: Implementation Guide*; Publications Office of the European Union: Luxembourg, 2020. Available online: https://data.europa.eu/doi/10.2760/936769 (accessed on 21 February 2025).
5. Sitikhu, P.; Pahi, K.; Thapa, P.; Shakya, S. A comparison of semantic similarity methods for maximum human interpretability. In *2019 Artificial Intelligence for Transforming Business and Society (AITB)*; IEEE: New York, NY, USA, 2019; pp. 1–4.
6. De Vries, D.A.; Piotrowski, J.T.; De Vreese, C. Developing the DigIQ: A measure of digital competence. *PLoS ONE* **2025**, *20*, e0322995. [CrossRef]
7. Ferrari, A. *DIGCOMP: A Framework for Developing and Understanding Digital Competence in Europe*; Publications Office of the European Union: Luxembourg, 2013. Available online: https://data.europa.eu/doi/10.2788/52966 (accessed on 30 January 2025).
8. Vuorikari, R.; Punie, Y.; Gomez, S.C.; Van Den Brande, G. *DigComp 2.0: The Digital Competence Framework for Citizens. Update Phase 1: The Conceptual Reference Model*; Joint Research Centre (Seville Site): Sevilla, Spain, 2016.
9. Carretero Gomez, S.; Vuorikari, R.; Punie, Y. *DigComp 2.1: The Digital Competence Framework for Citizens with Eight Proficiency Levels and Examples of Use*; Publications Office of the European Union: Luxembourg, 2017. Available online: https://data.europa.eu/doi/10.2760/38842 (accessed on 28 February 2025).
10. Redmond, B. Liderazgo de equipo compartido de un programa online. *Rev. Esp. Pedagog.* **2020**, *78*, 89–100. [CrossRef]
11. Redecker, C. *European Framework for the Digital Competence of Educators: DigCompEdu*; Publications Office of the European Union: Luxembourg, 2017. Available online: https://data.europa.eu/doi/10.2760/178382 (accessed on 28 February 2025).
12. Mattar, J.; Santos, C.C.; Cuque, L.M. Analysis and comparison of international digital competence frameworks for education. *Educ. Sci.* **2022**, *12*, 932. [CrossRef]
13. Banoy-Suarez, W.; González-Reyes, R.A. Analysis of Frameworks for Digital Skills Training for Secondary School Teachers: A Systematic Review. *TEM J.* **2024**, *13*, 1038. [CrossRef]
14. Vuorikari, R. DigComp Helping Shape the Education Ecosystem in Europe. In *Shaping the Digital Transformation of the Education Ecosystem in Europe*; Väljataga, T., Laanpere, M., Eds.; Springer International Publishing: Cham, Switzerland, 2022; pp. 137–143.
15. Caena, F.; Redecker, C. Aligning teacher competence frameworks to 21st century challenges: The case for the European Digital Competence Framework for Educators (*Digcompedu*). *Eur. J. Educ.* **2019**, *54*, 356–369. [CrossRef]
16. Cabero-Almenara, J.; Gutiérrez-Castillo, J.-J.; Palacios-Rodríguez, A.; Barroso-Osuna, J. Development of the Teacher Digital Competence Validation of DigCompEdu Check-In Questionnaire in the University Context of Andalusia (Spain). *Sustainability* **2020**, *12*, 6094. [CrossRef]
17. Builes, S.; Viñoles-Cosentino, V.; Esteve-Mon, F.M.; Sánchez-Tarazaga, L. Digital training in university teaching induction programmes in Spain: A comparative analysis based on DigComp and DigCompEdu. *Educ. XX1* **2024**, *27*, 37–64. [CrossRef]

18. Mattar, J.; Piovezan, M.B.; Souza, S.; Santos, C.C.; Santos, A.I.D. Apresentação crítica do Quadro Europeu de Competência Digital (DigComp) e modelos relacionados. *Res. Soc. Dev.* **2020**, *9*, e172943062. [CrossRef]
19. Mattar, J.; Ramos, D.K.; Lucas, M.R. DigComp-Based Digital competence Assessment Tools: Literature Review and Instrument Analysis. *Educ. Inf. Technol.* **2022**, *27*, 10843–10867. [CrossRef]
20. González-Mujico, F.d.L. Measuring student and educator digital competence beyond self-assessment: Developing and validating two rubric-based frameworks. *Educ. Inf. Technol.* **2024**, *29*, 13299–13324.
21. Garman, A.N.; Erwin, T.S.; Garman, T.R.; Kim, D.H. Developing competency frameworks using natural language processing: An exploratory study. *J. Competency-Based Educ.* **2021**, *6*, e01256. [CrossRef]
22. Zare, D.; Fernandez-Sanz, L.; Pospelova, V.; López-Baldominos, I. NLP and Text Mining for Enriching IT Professional Skills Frameworks. *Appl. Sci.* **2025**, *15*, 9634. [CrossRef]
23. Zare, D.; Fernandez-Sanz, L.; Pospelova, V.; Lopez-Baldominos, I.; Rizun, N.; Revina, A. Semantic mapping of soft skills from skillsmatch framework to the standard EN 16234-1 using transformer-based NLP. *Comput. Stand. Interfaces* **2026**, *98*, 104162. [CrossRef]
24. Fareri, S.; Melluso, N.; Chiarello, F.; Fantoni, G. SkillNER: Mining and Mapping Soft Skills from any Text. *Expert Syst. Appl.* **2021**, *184*, 115544. [CrossRef]
25. Khaouja, I.; Kassou, I.; Ghogho, M. A Survey on Skill Identification From Online Job Ads. *IEEE Access* **2021**, *9*, 118134–118153. [CrossRef]
26. Kang, P.S.; Enstroem, R.; Bhawna, B.; Bennett, O. A text mining study of competencies in modern supply chain management with skillset mapping. *Supply Chain Anal.* **2025**, *10*, 100117. [CrossRef]
27. Roh, D.; Yoo, J.; Ok, H. Mapping digital literacy in language education: A comparative analysis of national curriculum standards using text as data approach. *Educ. Inf. Technol.* **2025**, *30*, 6287–6313. [CrossRef]
28. Delcker, J. Digitalisation in the curricula of vocational schools: Text mining as an instrument of curricula analysis. *Technol. Knowl. Learn.* **2023**, *28*, 999–1014.
29. Aşıcı, F.; Temel, H. Analysis of secondary school mathematics textbooks in the context of digital competence. *J. Educ. Technol. Online Learn.* **2023**, *6*, 885–910. [CrossRef]
30. Ma, H.; Ismail, L. Bibliometric analysis and systematic review of digital competence in education. *Humanit. Soc. Sci. Commun.* **2025**, *12*, 185. [CrossRef]
31. Riko, R.; Joko, S.; Hasrini, S. Bridging the Competency Gap: Aligning Digital Competency Framework 2.2 for Citizens with Public Sector Digital Competencies for Industry 4.0. In *2025 IEEE International Conference on Industry 4.0, Artificial Intelligence, and Communications Technology (IAICT)*; IEEE: New York, NY, USA, 2025; pp. 242–248.
32. Van Audenhove, L.; Vermeire, L.; Van den Broeck, W.; Demeulenaere, A. Data literacy in the new EU DigComp 2.2 framework how DigComp defines competences on artificial intelligence, internet of things and data. *Inf. Learn. Sci.* **2024**, *125*, 406–436. [CrossRef]
33. Panzaru, C.; Grama, A. Towards explicit soft skills labelling in ESCO through semantic NLP analysis. *J. Labour Mark. Res.* **2025**, *59*, 18. [CrossRef] [PubMed]
34. Xun, L.S.; Gottipati, S.; Shankararaman, V. Text-mining approach for verifying alignment of information systems curriculum with industry skills. In *Proceedings of the 2015 International Conference on Information Technology Based Higher Education and Training (ITHET), Costa da Caparica, Portugal, 11–13 June 2015*; IEEE: Lisbon, Portugal, 2015; pp. 1–6.
35. Clavié, B.; Soulié, G. Large Language Models as Batteries-Included Zero-Shot ESCO Skills Matchers. *arXiv* **2023**, arXiv:2307.03539v1. [CrossRef]
36. Osipovskaya, E.; Zare, D.; Fernández-Sanz, L.; Tasi, P. Leveraging Digital Techniques for the Analysis and Study of Digital Skills. In *EDULEARN25 Proceedings*; IATED: Valencia, Spain, 2025; pp. 3776–3784.
37. Gribova, V.; Shalfeeva, E. SMART Standards for Industry. In *Proceedings of the Eighth International Scientific Conference "Intelligent Information Technologies for Industry" (IITI'24), Volume 1*; Kovalev, S., Kotenko, I., Sukhanov, A., Li, Y., Li, Y., Eds.; Springer Nature: Cham, Switzerland, 2024; pp. 70–82.
38. Narayanan, A.; Lechevalier, D.; Morris, K.C.; Rachuri, S. Communicating standards through structured terminology. *Comput. Stand. Interfaces* **2015**, *40*, 34–41. [CrossRef]
39. Ellis, W. Management of Technical Terminology in Standards Organizations. In *Standardization of Technical Terminology: Principles and Practices (Second Volume)*; Strehlow, R., Ed.; ASTM International: West Conshohocken, PA, USA, 1988; Volume STP991-EB.
40. Dunn, A.; Dagdelen, J.; Walker, N.; Lee, S.; Rosen, A.S.; Ceder, G.; Persson, K.; Jain, A. Structured information extraction from complex scientific text with fine-tuned large language models. *arXiv* **2022**, arXiv:2212.05238. [CrossRef]
41. Li, H.; Mao, H.; Wang, J. Part-of-Speech Tagging with Rule-Based Data Preprocessing and Transformer. *Electronics* **2021**, *11*, 56. [CrossRef]
42. Wan, Q.; Xu, X.; Han, J. A dimensionality reduction method for large-scale group decision-making using TF-IDF feature similarity and information loss entropy. *Appl. Soft Comput.* **2024**, *150*, 111039. [CrossRef]

43. Chiche, A.; Yitagesu, B. Part of speech tagging: A systematic review of deep learning and machine learning approaches. *J. Big Data* **2022**, *9*, 10. [CrossRef]
44. Krathwohl, D.R. A Revision of Bloom's Taxonomy: An Overview. *Theory Into Pract.* **2002**, *41*, 212–218. [CrossRef] [PubMed]

Disclaimer/Publisher's Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.