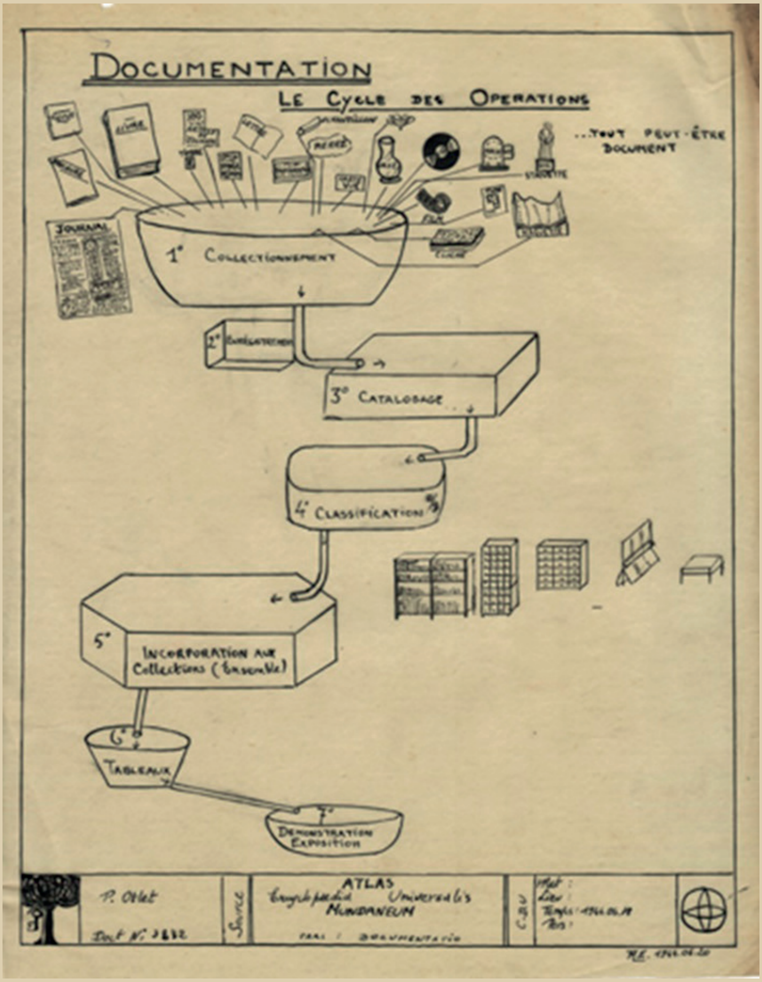
AIDAinformazioni

RIVISTA SEMESTRALE DI SCIENZE DELL'INFORMAZIONE

NUMERO 1-2
ANNO 43
GENNAIO-GIUGNO 2025



CACUCCI EDITORE
BARI

AIDAinformazioni

RIVISTA SEMESTRALE DI SCIENZE DELL'INFORMAZIONE

Fondata nel 1983 da Paolo Bisogno

Editrice: Cacucci Editore S.a.s.
Via D. Nicolai, 39 – 70122 Bari (BA)
www.cacuccieditore.it
e-mail: riviste@cacuccieditore.it
Telefono 080/5214220

AIDAinformazioni

Rivista semestrale di Scienze dell'Informazione

«AIDAinformazioni» è una rivista scientifica che pubblica articoli inerenti alle Scienze dell'Informazione, alla Documentazione, all'Archivistica, alla Gestione Documentale e all'Organizzazione della Conoscenza ma amplia i suoi confini in ulteriori campi di ricerca affini quali la Terminologia, la Linguistica Computazionale, la Statistica Testuale, ecc. È stata fondata nel 1983 quale rivista ufficiale dell'Associazione Italiana di Documentazione Avanzata e nel febbraio 2014 è stata acquisita dal Laboratorio di Documentazione dell'Università della Calabria. La rivista si propone di promuovere studi interdisciplinari oltre che la cooperazione e il dialogo tra profili professionali aventi competenze diverse, ma interdipendenti. I contributi pubblicati affrontano questioni teoriche, metodologie adottate e risultati ottenuti in attività di ricerca o progettuali, definizione di approcci metodologici originali e innovativi, analisi dello stato dell'arte, ecc.

«AIDAinformazioni» è riconosciuta dall'ANVUR come rivista di Classe A per l'Area 11 – Gruppo Scientifico Disciplinare 11/HIST-04 – Scienze del libro, del documento e storico-religiose e come rivista scientifica per le Aree 10 – Scienze dell'antichità, filologico-letterarie e storico-artistiche; 11 – Scienze storiche, filosofiche, pedagogiche e psicologiche; 12 – Scienze giuridiche; 14 – Scienze politiche e sociali. È anche annoverata dall'ARES (Agence d'évaluation de la recherche et de l'enseignement supérieur) tra le riviste scientifiche dell'ambito delle Scienze dell'Informazione e della Comunicazione. La rivista è, inoltre, indicizzata in: ACNP – Catalogo Italiano dei Periodici; BASE –Bielefeld Academic Search Engine; ERIH PLUS – European Reference Index for the Humanities and Social Sciences – EZB – Elektronische Zeitschriftenbibliothek – Universitätsbibliothek Regensburg; Gateway Bayern; KVK – Karlsruhe Virtual Catalog; The Library Catalog of Georgetown University; SBN – Italian union catalogue; Ulrich's; Union Catalog of Canada; LIBRIS – Union Catalogue of Swedish Libraries; Worldcat.

I contributi sono valutati seguendo il sistema del *double blind peer review*: gli articoli ricevuti sono inviati in forma anonima a due referee, selezionati sulla base della loro comprovata esperienza nei topics specifici del contributo in valutazione.

AIDAinformazioni

Anno 43

N. 1-2 – gennaio-giugno 2025

CACUCCI EDITORE

BARI

Via Nicolai, 39 – 70122 Bari – Tel. 080/5214220
http://www.cacuccieditore.it e-mail: info@cacucci.it

Sommario

Contributi

Rubriche

Contributi

AIDAinformazioni
ISSN 1121–0095
ISBN 979-12-5965-549-3
DOI 10.57574/596554932
pag. 23-44 (gennaio-giugno 2025)

Enhancing technoscientific heritage and introducing concepts into culture in the age of data, semantic web, and Artificial Intelligence

The ITinHeritage project

Caroline Djambian*

Abstract: In Semantic Web age, enhancing heritage means bringing collections beyond the confines of the museum. But this ethical mission of transmitting culture and knowledge to the general public is made complex by the massive and heterogeneous data that museums now have to manage. As data are now the new collections of museums, they are the starting point for the ITinHeritage project. It focuses on Information Technology (IT) heritage, the preservation of which is of general interest, and raises the question of "How to pass on complex scientific and technical concepts embodied in artefacts not mediators in themselves?" Backed by European science and IT museums, we develop innovative Digital Humanities approaches, combining terminology, Linked Open Data (LOD), and Artificial Intelligence (AI) technologies. In this article we set out the context of our questioning and present the basis of the answer we wish to provide, i.e. the onto-terminological work.

Keywords: Digital Humanities, Scientific and technical heritage, Ontology, Knowledge graph, Linked Open Data (LOD).

1. Introduction

The ITinHeritage project was born of a meeting, as a terminologist and researcher in organization of scientific and technical knowledge in its digital mediation, with members of the Association pour un Conservatoire de l'Informatique et de la Télématique (ACONIT) in Grenoble (France). Grenoble is the birthplace of French computer science, and the Conservatory owns one of the largest collections in Europe. If Information Technology (IT) is the emblem and the witness of contemporary human technoscientific evolution, and the heart of our current society, the more I got to know this museum space, the more I realised how, paradoxically, little IT heritage is known and understood by the public and authorities, and therefor how endangered it was. Further investigations revealed that the difficulties encountered by the

* GRESEC Laboratory (Groupe de recherche sur les enjeux de la communication), Grenoble Alpes University, Grenoble, France. caroline.djambian@univ-grenoble-alpes.fr. ORCID: 0009-0007-3625-305X.

Conservatory were general to IT museums. That's why I decided to create a network of several European museums around a common desire to promote and pass on this heritage. This gave me a fantastic research area with collections of thousands of artefacts. But how, then, preserve this heritage and pass on complex scientific and technical concepts embodied in artefacts that are not mediators in themselves, both to scientists and to the general public?

To answer this question, my initial idea was to use IT technologies to serve their own heritage. So, I began by basing my study on current digital developments embodied in major semantic web projects, which offer an unprecedented opportunity to enhance and disseminate heritage and the knowledge it conveys to the public. These Digital Humanities projects from the Galleries, Libraries, Archives, Museums (GLAM) movement, are based above all on the metadata's standardization work (here, inventories of museum collections), in order to open them up. For example, we can mention Europeana (Doerr et al. 2010), the Société des Musées du Québec, the Royal Rijksmuseum of the Netherlands around the works of the great masters of Dutch painting (Dijkshoorn et al. 2018), JocondeLab for the museums of France (Juanals and Minel 2016), the Smithsonian American Art Museum (Szekely et al. 2013), the Amsterdam Museum (De Boer et al. 2013) and the Getty in Los Angeles, the Bibliothèque Nationale de France with data.Bnf (Simon et al. 2014), Biblissima (Gehrke et al. 2015).

But while this represent a major task, it cannot be an end in itself, because making data easier to consult is not enough to guarantee the museum's mission of transferring culture and knowledge. If data is the starting point, the appropriation of knowledge is our objective, and it is through language that it can be captured. That's why we will focus here on the terminological work that founds the ITinHeritage project, carried out with the help of a multidisciplinary team from the fields of information and communication sciences, computer sciences, linguistics and history of science.

After having presented the particularities of IT's technoscientific heritage, we will look in this paper at: how this heritage, like others, is tending towards the immateriality of data; and how the latter is becoming the raw material for a revisited corpus linguistics approach, placing language at the heart of the transfer and appropriation of knowledge that we are aiming for. On this basis, we will explain why, given the complexity and variability of technoscientific language, the semasiological approach cannot be sufficient, but rather the first stage of a terminological work. Finally, we will explain our method for building an onto-terminology of IT heritage, and how we complement it by Artificial Intelligence (AI) technologies, in order to semantically link museums' open data, while representing technology in its dynamic evolution.

2. An evolving technoscientific subject for study: the Information Technology (IT) heritage

Technosciences, these complex systems that combine technology and science, are the foundation of our contemporary world. They are developing alongside our humanity, transforming it on an unprecedented scale and at an unprecedented speed. Information Technology (IT) is the very embodiment of technosciences and their relationship to our society. They are defined by UNESCO (2009, 121) as «a diverse set of technological tools and resources used to transmit, store, create, share or exchange information. These technological tools and resources include computers, the Internet (websites, blogs and emails), live broadcasting technologies (radio, television and webcasting), recorded broadcasting technologies (podcasting, audio and video players, and storage devices) and telephony (fixed or mobile, satellite, visio/video-conferencing, etc.)». But even though Information Technologies are pervading our humanity, their rapid and massive evolution has not left room for their in-depth and distanced study, for their definition and for their patrimonalization.

Yet studying this heritage can give us a better understanding of our contemporary world and how it has changed since the Second World War. That's what we are aiming for in the ITinHeritage research project, through an heritage-based and interdisciplinary approach funded in Digital Humanities (Djambian et al. 2024a). Our project starts from European scientific and IT museums that long ago began the painstaking work of inheriting, conserving and promoting this heritage. The London Science Museum (London, UK) and the Heinz Nixdorf MuseumsForum HNF (Paderborn, Germany), the Association pour un Conservatoire de l'Informatique et de la Télématique (ACONIT) (Grenoble, France), the NAM-IP computer museum (Namur, Belgium), the Museo degli strumenti per il calcolo (Pisa, Italy) and the HomeComputerMuseum (Helmond, Netherlands) are our partners.

The aim of our study is therefore to define and circumscribe this labile field, through the various forms that its heritage can take. The first form of this heritage is the one that is most obvious at first glance: its physical and digital objects that are its explicit expressions. This includes objects and their documentation, their digital duplicate, the software that represents 1/3 of certain collections, and above all, the metadata of all these artefacts. The growing proportion of data makes it the new collections of science. But it is the tacit expression of this heritage, and the knowledge it encompasses, that we would like to focus on. Much more difficult to grasp, because it is created in the act of experiencing the world, it has been little studied and even less considered as heritage. It echoes the empeiria of Kant (1997) or the metis of Aristotle (1986), whereas the exoteric (Jacob 2001) or pure knowledge of Kant is concretized in writing or in object. Gathering and transmitting this tacit, esoteric knowledge, is a real challenge because, to be captured, empeiria needs to be shaped into gestures and discourse. We are therefore placing the specialized language at the heart of the ITinHeritage project by combining complementary and innovative approaches aimed at the interaction between terminology, in its conceptual nature, and Artificial Intelligence (AI).

3. New museums' collections and new bodies of work: the data

Data are not only the new artefacts of heritage, they are also the new corpora for researchers (Djambian et al. 2024b). As with many research projects, our terminology work begins with the compilation of a corpus. To build it, we have opted for an approach based more on the information and communication sciences than on linguistics, focusing not on the volume of the corpus but on its high density of specialized terms and concepts, as manifestations of the tacit knowledge that we want to enhance. To this end, we have built a Knowledge Graph that gathers the metadata from our partners museums' inventories of collections. Knowledge Graphs are part of the Semantic Web paradigm, which enable to open and link data with other data from around the world and to easily integrate new data from the web, in a reciprocal enrichment process. To build the ITinHeritage Knowledge Graph we used the Jena environment. We first collected and harmonized the metadata of the various museums, as each one uses its own formats (XML, PDF, ...) and categorises knowledge according to its own references. This harmonization followed the European open data FAIR principles (Findable, Accessible, Interoperable, Reusable). Then we converted metadata to CSV (Comma-Separated Values) and RDF (Resource Description Framework) standards so that it could be opened and linked on the web (Linked Open Data, LOD). But, in order to move from the web of data to the semantic web, it is essential to structure the Knowledge Graph using an ontology. So, we built on the Protégé environment a first-level ontology based on the CIDOC-CRM (International Committee for Documentation of the International Council of Museums, Conceptual Reference Model) (Bekiari 2021) and the EDM (Europeana Data Model) (Doerr et al. 2010) (Fig. 1), two meta-models that are benchmarks in the heritage field.

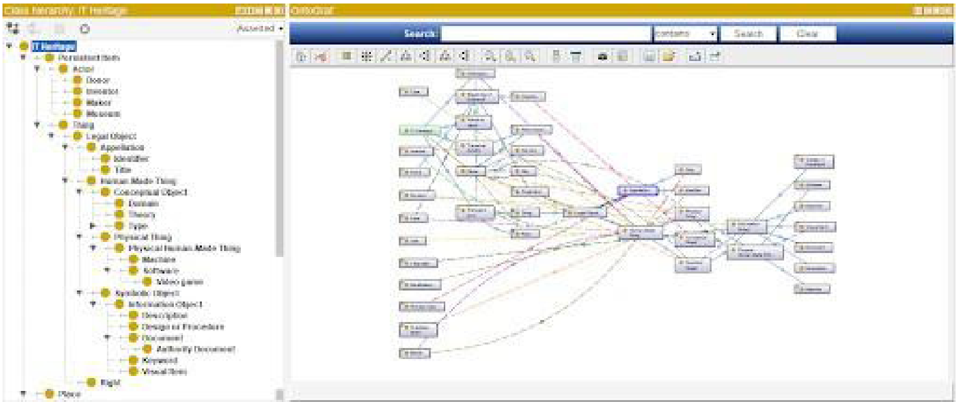


Figure 1: View of he first-level ITinHeritage ontology.

The metadata of 25,500 artifacts are now contained in our Knowledge Graph. The primary aim is to open up museum collections beyond their walls, so that this data can be made available to the public at large, in a spirit of open science. But for us, this Knowledge Graph represent above all our corpus, packed with terms and concepts in the field, written by experts. It represents the new types of corpora resulting from Big Data, which linguists will now have to encounter. In order to define the tacit knowledge of the field, this multilingual corpus (French, English, Italian, German) allows us to study the discourses that express it. This is the starting point for our onto-terminological work based on ISO standards (5078, 704 and 1087).

4. How can technoscientific concepts be made appropriate in the face of the variety and technicality of language?

If we focus on language, we must take it in all its complexity. The first barrier to dealing with technoscientific language is its complexity. Bertrand Russell (1956) asserted that in an ideal language, a word would designate an object, and each object would have its own meaning. However, the complexity of language was already noted in antiquity by Democritus, who observed that a word can designate several objects, that the same object can be designated by several words, that words designating objects vary over time, and finally, that the links between a word and an object have various motivations (Diels 1903). This is why, in languages as complex as technosciences, it is necessary to come to a conceptualization of a world whose elements are designated by unambiguous names (Roche 2007). This step is not an option when we want to transmit specialized knowledge to a wide audience.

To be able to pass on technoscientific knowledge to the layman, the first step is to explain the meaning of the names of the concepts, i.e. the specific terms of the field. It is the notional dimension that makes it possible to cover the various dimensions of a domain and achieve the objective of popularization of its knowledge. To this end, terminology work cannot be carried out without the involvement of the holders of these concepts, as in our case, experts and historians of IT. That's why our work is carried out with the support of Pr. Marie Gevers and Christophe Ponsard (Namur University) from the NAM-IP museum, Xavier Hiron from the ACONIT conservatory, and other experts from the HomeComputerMuseum and the Museo degli Strumenti per il Calcolo).

As we shall see in detail below, we need to integrate a meta-linguistic discourse, and textual and semantic strategies, that will compensate for the accumulation of specialized terms in specialized discourse. In order to make specialized terms, i.e. the names of concepts, appropriable, the descriptive strategy to integrate technoscientific concepts into the culture of the individual receiver must be based on the differentiation of notions (Gaudin 1996).

This is because technoscientific language in itself constitutes an obstacle to access to specialist knowledge, just as much as the lack of mediation of the scientific and technical objects it names. It is the concept that can be appropriated rather than the terms that designate it, even if Putnam (1984) demonstrates that we can manipulate specialized languages without mastering their notions. This is because acquiring a lexical system is not enough, and remains a barrier to the uninitiated because of its opacity. Efforts should therefore be focused on the appropriation of a conceptual system, i.e. on terminology work in addition to linguistic work.

As we said earlier, understanding a new term is based on comparing it with, or differentiating it from a pre-existing notional network. However, the difficulty of integrating new knowledge persists when the existing notional network has no link, no common references, to the new object and/or term. As presented in previous works (Djambian et al. 2024a) and based on those of Gaudin (1996), we can then turn towards two possibilities: the exploitation of semantic-syntactic relations and lexical relations.

Firstly, semantic-syntactic relations can be analyzed and exploited to express complex knowledge. We refer to relationships such as "typical object", "typical action" and "typical agent" (Lerat 1987 and 1988), and "typical application" (Gaudin 1996, 17). These relations list the typical collocations of a lexical unit, making it possible to provide a detailed description of the use made of the object named, or in some way, to substitute for an experience that we could have of the object in the real world.

This is particularly well suited to our ITinHeritage project, with metadata describing the artefacts of partners museums. Here is an example of the description of the artifact named “Tabulator BS 120” from the ACONIT Conservatory (Grenoble, France):

> La tabulatrice Bull BS 120 se compose d’un lecteur de cartes (traitant 150 cartes par minutes), d’une imprimante (cadence de 150 lignes par minute sur une largeur de 92 colonnes), d’une perforatrice de cartes (débitant soixante-quinze cartes par minute). Elle dispose d’un calculateur mécanique qui lui permet d’exécuter les quatre opérations arithmétiques, des opérations logiques et de mémoriser des informations. Elle offre un système de programmation amovible : le ”tableau de connexion”, qui était spécialement câblé pour chaque traitement. Sa technologie est entièrement électromécanique.

- Typical action: lire, additionner et imprimer
- Typical object: caractères
- Typical agent: calculateur mécanique, lecteur de cartes, imprimante, perforatrice de cartes (ateliers mécanographiques à cartes perforées)
- Typical application: calcul électromécanique

Semantic-syntactic relations are proving to be a valuable tool for translating technoscientific discourses and gradually make the work of textual analysis moving towards that of knowledge representation. Their interest lies in the contextualization of the object in its uses. This helps to gain understanding, even if it is still insufficient to acquire knowledge. The path towards the concept to be understood is thus supported by a description of the object’s properties, echoing some characteristics of the concept. We will present below, another method of working with terminology along these lines.

However, we clearly see in this example that relying only on semantic-syntactic relationships is insufficient to access meaning in the face of the complexity and density of technoscientific language. A translation work related to the recipient’s references remains necessary, but excessively burdensome if carried out manually, for instance, for the construction of definitions. We can take the example of the typical application “electromechanical calculation” which can be presented as a “precursor of computer programming”, and provides the non-expert with the most commonly accepted reference of “computer programming”. It is thus possible to use the inclusions of technoscientific languages in everyday language, even if not all fields are represented in the same proportions, because a specialized language is not to be understood as a whole with defined boundaries, and because it abounds in obstacles to the appropriation of knowledge. In our field of Information Technologies, which can serve as an example, we can observe complex terms (mechanographic workshops with punched cards), eponyms (Turing machine, Moore’s law), acronyms complicated by version numbers (IBM 1130), etc.

However, contemporary technosciences are characterized by both a great technicality and a high degree of socialization. Their objects shape our daily lives, especially in the realm of Information Technologies. Consequently, many of their terms sketch out certain notions within us, such as a "keyboard" or a "computer mouse". As demonstrated by Gaudin and Lerat, as well as based on empiricist philosophical theories (Aristotle, Condillac), it is through the application made around these objects, that is to say, the experience that humans have of these objects, that meaning can be constructed towards more complex notions and that names can be attached to these objects. The common language and the specialized language overlap and enrich each other in the wake of the adoption of technologies by the general public.

However, another significant obstacle lies in the linguistic diversities within the same field. These diversities can be geographical (diatopic), temporal according to technological developments (diachronic), or community-based (diastratic). Continuing with Information Technologies, a "first-generation calculator" for the amateur is referred to as a "mechanical calculator" by a historian of computing, or as an "antique calculator" by a general historian. One of the objects that this concept covers is the "abacus", as the "first digital tablet". It is obvious that, for the public of the 21st century, a "digital tablet" refers to a completely different object, i.e. a modern digital pad (laptop). But the object designated in the IT heritage dates back to Antiquity and consists of clay balls and tokens, from the romans "calculi". These abacuses were used for arithmetic until 7000 BC, and evolved into the instrument with rows of movable pieces, better known from 500 BC as the "abacus". It is also interesting to note that English does not distinguish the evolution from the abacus to the "boulier", which for French are two distinct objects (abaque/boulier). These linguistic differences within the same field reflect distinct representations of the world, which are themselves derived from different experiences of it: «... we only manipulate reality through the representations we have of it» (Roche 2005, 50).

In the field of IT, as in all technosciences, especially those that are socially established and characterized by rapid or abundant technological evolution, we can observe these linguistic divergences. Another example that we cite in our previous works (Djambian et al. 2024a) which can illustrate these divergent representations of the world within the same domain, due to different experiences of it, is that of the "Baby" as amateurs name it. The museums designate it as the "Manchester Baby" in reference to its place of construction, by the more comprehensive formulation of "Manchester Small-Scale Experimental Machine" (London Science Museum), because the "Small-Scale Experimental Machine" (SSEM). It was the world's first von Neumann architecture machine built at Manchester's Victoria University in 1948, known by the general public as the first "computer". But, according to historian of computing, mainstream "computers" are "stored-program computers", in the strict sense of the term, i.e. von Neumann-type computers, which have a very precise meaning: they are electronic calculating machines with a central memory large enough to hold the program being executed, as well as the data. This definition makes the EDVAC (Electronic Discrete Variable Automatic Computer) the first electronic stored program computer, developed by the Pennsylvania University for the United States Army. So, we can see the extent to which the variety of languages in use reflects different representations of the world and its objects, and even impacts the vision of the history of a field, and can reflect underlying political and economic hegemonies and socio-cultures.

These linguistic differences within the same field can only lead to a conceptual confusion. However, in the relationship between everyday language and specialized sub-languages (diastraties), words remain the sole foundation that allows us to access the concepts they designate, especially nowadays where access to knowledge primarily occurs through digital or, at least, formal mediation. Regarding IT heritage, we find designations of concepts in the texts composed of the metadata that describe the collections of partner museums and serve as the basis for our terminological work. Given the variety of languages used in the field, this choice is based on the fact that these texts are written by experts in the field, in a language that is already sufficiently controlled to achieve consensus and understanding by different audiences.

Through a linguistic analysis, we are able to extract phrases that can maintain lexical relationships. Generic or partitive lexical relationships play an essential role in the appropriation of knowledge and prove to be more effective for terminological work. According to ISO 704 standard, they form its foundation and allow for a more immediate positioning of the term within its conceptual context. They place it within a whole. Lexical relationships are the second opportunity provided for the integration of new terms and knowledge, and they are strongly represented in our corpus, as evidenced by the description presented above of the “Tabulator BS 120”.

5. The semasiological approach underlying terminological work

Our previous works (Djambian et al. 2024a) have shown us the usefulness of an analysis of linguistic uses as a first step to identify the names of concepts and their meanings. It is therefore on the basis of a semasiological approach that we support our terminological work, starting from our Knowledge Graph which gathers the metadata of the partners museums’ inventory collections. The construction of our corpus has followed the text selection criteria of ISO 5078 (ISO 2025). This standard has also guided our semi-automatic lexical extraction work, where human intervention has always complemented automatic extraction. To this end, we have combined statistical and linguistic techniques. Our work initially focused on the largest corpus, which is the French corpus of the ACONIT Conservatory, selecting the fields that are richest in terms of texts: Description, Name, Model, and Use. To carry out the automatic lexical extraction, we turned to the TermoStat (Drouin 2003) and Sketch Engine (Kilgarriff et al. 2014) extraction tools. We relied on Python's Natural Language Toolkit (NLTK) too, composed of libraries and programs for Natural Language Processing (NLP), supporting classification, tokenization, stemming, tagging, parsing, and semantic reasoning functionalities (Bird et al. 2009). Indeed, programming languages are now considered relevant alternatives to traditional extraction tools when facing corpora derived from big data (Anthony 2021).

Our analysis was first based on the TermoStat extraction tool. It has made it possible to highlight simple and complex terms among lexical categories such as nouns, adjectives and verbs. It should be noted that TermoStat bases its analysis on a method of contrast between specialized and non-specialized corpora. The terms we are looking for are the most specific to the field and therefore those which have a lower frequency in the general reference corpus. However, TermoStat has shown real limitations regarding the terms specific to technosciences such as IT: for example, it is impossible to extract acronyms or terms starting with a capital letter followed by a series of numbers. The extraction tools prove to be poorly suited for these languages. These cases, related in the field of IT, to machines' names and their models or manufacturers, are nonetheless of critical importance in the domain (e.g., Gamma 30, IBM 1130, ...). Another problem is that TermoStat understands the Gamma 3 electronic calculator, marketed by Bull in 1952, as a verb. So, this first phase of extractions done by Termostat, which we had hoped would be more successful with the help of Python's NLKT library, proved unsatisfactory. The level of noise required an excessively demanding manual intervention.

Finally, Sketch Engine is the tool that has demonstrated the best results. For the lexical extraction, we considered with particular attention the multi-words terms results (complex terms) such as "mémoire à tores magnétiques", which are the most frequent lexical pattern in IT and the most meaningful. The first single-word terms obtained, such as "micro-ordinateur", "disquette", "macintosh", "microprocesseur", "azerty", "hewlett-packard", "powerbook", "olivetti", and "alphanumérique" are mainly generic terms, often the ones most commonly used in everyday language, and therefore the most familiar to the public, but less significant in the field. They are not part of the specialized expert language, unlike the complex terms, acronyms or terms beginning with an uppercase letter followed by a series of numbers. For these, the Corpus Query Language (CQL) of the Concordance section of Sketch Engine was used because it allows the analysis in various possible contexts, extracts, and analyzes more complex grammatical patterns.

Some observations that we detailed in precedents works (Rossi et al. to be published), have been made based on this first linguistic analysis. The verbs, which are very present in our corpus, often serve as markers of partitive relationships (Lefeuvre and Condamines 2017), such as “to possess”, “to contain”, “to include”, and “to involve”. They indicate components of machines or their technical functioning. The adjectives, for their part, mainly refer to the shape or function of the machines. If the Description field pertains to a museum mediation strategy centered on a detailed description of the object, the Usage field engages in a narrative mediation with a complex structure that, beyond the artifact, describes the practices or actors related to the history of the object. The verbs here are more often focused on the relationship between actions or processes: “to enable”, “to be able”, “to serve”, “to act”, and “to function”.

We have worked on the same fields within the English language corpus formed by the metadata from the London Science Museum and the HNF for the purposes of unbiased comparative study. The lexical extraction work on the “Description” field shows, contrary to the French corpus, a focus on artefacts but also on their contextualization, a mediation strategy that we had observed in the “Use” field in French. This significant difference in mediations reflects a difference in the representations of the world, which is also observable in the Italian corpus from the Museo degli Strumenti per il Calcolo in Pisa. The language practices among these different cultures also reflect a distinct relationship to objects in their patrimonialization. Here, the Adj+Noun model is the most productive, with examples such as “personal computer”, “video game”, or “electronic calculator”. The most frequent verbs focus on concrete actions: “to manufacture”, “to build”, “to work”, although one can also find a significant number of occurrences for verbs such as “to include”, which more marginally, indicate partitive relationships. At this point, the semasiological analysis allows us to highlight some interesting trends within our corpora: the prevalence of a morphological pattern for the creation of terms in the domain, as well as the presence of markers of conceptual relationships.

6. About the need to use socially and historically established consensual representations: the necessary completion by the onomasiological approach

However, the conclusion reached at the end of this semasiological work is that it highlights the extreme dependence of the lexicons extracted with the initial corpora, and it does not resolve, quite the contrary, the linguistic variety and complexity of the domain. It can, in no way, in its current state, lead to a consensual conceptualization. In addition, what should we do with these countless lexicons? Our previous experiences have shown us the difficulty to build up a semantic network based on this semasiological work (Djambian 2011). It poses the problem of finding only the variety of names of the domain's concepts in the texts and not the concepts. This study, important foundation for the maturation of terminological work, is limited to the analysis of words in their uses. The previous lexical analysis already shows the disparities that appear from one language to another, but more simply, from one museum collection manager (the metadata writer) to another.

As we said above, one cannot limit oneself to studying how a field is named without being interested in the conceptualizations that underlie these linguistic usages. To compensate for linguistic variations and bring out a common meaning, it is advisable to take an interest in the extralinguistic part of terminology work, centered on the relationship of the concept to the object, and not just in the relationship between a term and its signified, as in the linguistic study we presented in the semasiological phase. The meaning of a word must be independent of usage and consensually standardized within a community that refers to a conceptualization of the world. Thus, it must be remembered that all terminological work should be based on concepts and not on terms (Felber 1987). Terms serve as a point of entry for the terminologist, as a denomination of a concept (Roche 2005). They represent an expression of a conceptual system that, in turn, reflects the way we perceive the objects in our world. The purpose of terminological work is then to establish a consensual and standardized conceptual system.

The onomasiological approach allows for a more comprehensive inclusion of linguistic specificities and variations, as well as the representation of the conceptual system. This formalization stage, conducted with the assistance of field experts, is based on the convention in the Latin sense of *foedus*, which captures the full dimension of the societal and historical context in which a concept is born and evolves. As the concept is an eminently social construction and, to exist, must be validated within a given collective receptor: «The validity of solitary thought is principally dependent on the justification of linguistic statements within the effective community of argumentation. It is not possible for a single being to follow a rule and validate his thought within the framework of a "private language". Rather, thought is public in principle» (Apel 1987, 90).

But how the conventions of intercomprehension and the terminologization of words are founded? For Kripke names are socially elaborated too and are linked to the concept only by «our interaction with other speakers in the community, an interaction under which we are linked to the referent itself» (Kripke 1982, 82). For him, the definition of real-world objects involves substituting for «the idea of necessary and sufficient properties, the idea of a bundle of properties, only some of which must be satisfied in each particular case» (Kripke 1982, 116), in other words, a prototype or what Wittgenstein also calls a family resemblance (1961). The reference fixed by the act of naming is perpetuated through evolutions in time, by successive interactions in communication, bequeathing these properties which, through abstraction, become conceptual essential characteristics. Hilary Putnam (1990, 47) discusses reference as a social consensus extended to «the contribution of the environment».

For objects of the real world, the reference, which may be fixed arbitrarily and consensually by standards, is ephemeral and refers us to the role of the expert, since in the socially situated exchange of language, there is a distribution of roles and a "linguistic division of labour". Hilary Putnam argues that meaning does not emerge simply by being related to our pre-existing thoughts, but depends on our context and our interaction with it. Identification and validation within this socially established context are integral to the construction of an individual conceptual universe. Putting a concept into words is therefore fundamental because it links the concept to a set of variable linguistic usages, corresponding to practices and organizing a representation of the world, a logosphere, as «the mesh of meaning enveloping reality» (Lafont 1993, 292). It therefore establishes a direct link between language and the objects of the real world and encompasses individual linguistic varieties and consensual community norms.

While ontological construction follows onomasiological paths, the consideration and study of discourse (the semasiological approach) remains essential for understanding the processes by which concepts are constructed, constantly reformulated in discourse, in a historical dimension that texts make observable. «The history of concepts plays a part in understanding them» (Gaudin 1996, 609). The concept-discourse dynamic is, therefore, bilateral and iterative and is built in a given socio-cultural context. By studying discourse, it is essential to take account of this social and historical construction to retrace the genesis of the concept and carry out what Alain Rey (1990, 778) calls «the archaeology of conceptual constructs, ideological and scientific systems, or in an oversimplified term, the history of ideas».

Through a logical formalization, the ontological construction will reproduce this "logosphere" within which experts, who have a socio-linguistic responsibility in the transfer of knowledge, will be able to express prototypes bringing together bundles of essential characteristics that are sufficiently consensual to encompass the varieties socially established in the various communities and over time. The organization of concepts in the form of a notional system requires an inquiry into the essential characteristics that form the concepts. This formalization must designate the concepts beyond their usages to explicitly position them within the overall organization of the notional network. The ordering of concepts according to their hierarchical relationships and differences allows for the construction of a standardized vocabulary based on the generic or partitive lexical relations previously observed during the lexical extraction process.

For example: a “supercomputer”, whose synonym is “supercalculator”, from 1961 to now, has the essential characteristics of being designed to achieve the highest possible performance, particularly in terms of computing speed, and of being a mainframe computer. It is a kind of “second generation stored program computer”, characterized by the use of “transistors and batch systems”. The “supercomputer” has evolved to a “massively parallel stored program computer” from 1980, which characteristic is to use a large number of “computer processors” or separate “computers” to perform a set of coordinated calculations in parallel.

The dimension of hyperonymy (is a kind of) or meronymy (is composed of) is managed in the ontology by the subsumption relationship inherent in the Porphyry tree organization of classes (Porphyre 1984). The synonymy relationship is managed in class (concept) annotations by labels, using models such as SKOS (Simple Knowledge Organization System) (e.g. skos:altLabel). The semantic-syntactic relationships mentioned above can also be reproduced in a logical formalization as an ontology by integrating certain concepts (classes) and using description logic based on the subject-predicate-object triptych. The subject (concept or class) is linked to the object (instance) by a predicate (property or relation), which enables real sentences to be modeled (Fig. 2). For example, the object (instance) “punched card tabulator” is a “mechanographic machine”, which is a “calculation device” (subsumption relation). It was invented in 1887 (Date class) by Herman Hollerith (Inventor class), who founded the International Business Machines Corporation (IBM) (Constructor class), for the 1890 census (Use class) of the United States (Place class). It is composed of “60 counter dials”, “punched cards”, “a reading head” (Description class).

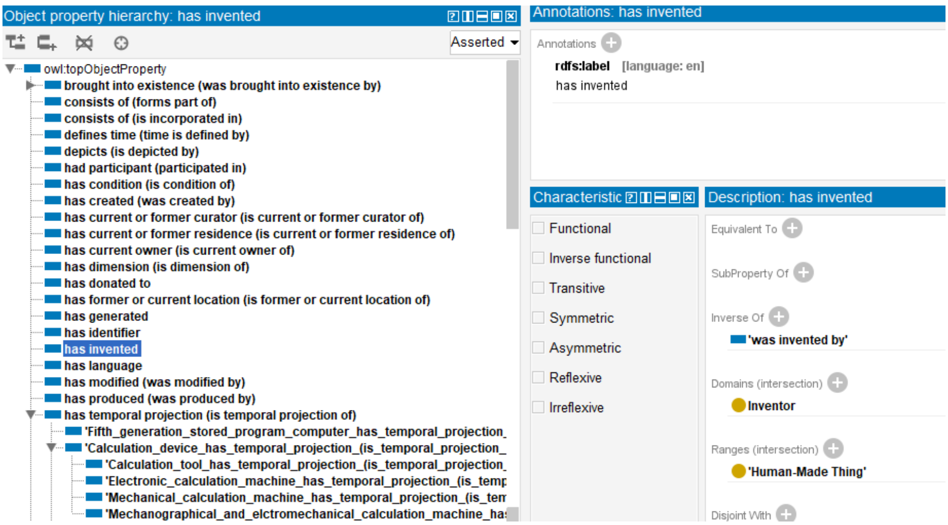


Figure 2: An example of the “has invited” property linking inventors to their inventions.

The construction of the IT domain's onto-terminology is facilitated by the use of, in a first time, the Protégé environment for the Knowledge Graph structuration, and in a second time, on the TEDI tool (Termino-ontological EDItor) (Roche and Papadopoulou 2019), specially designed to carry out terminological work according to the ISO 704 and 1087 standards. In particular, it allows to describe precisely the essential and distinctive characteristics of each concept, thus to define the ontology and analyze the conceptualization in the field, while generating at the same time, a dictionary, always built in accordance with the ISO standards mentioned above. In response to the study of the semantico-syntactic relations that we have described above, which are intended to contextualize a concept by highlighting the properties of an object that emerge from its uses, terminology work on TEDI enables us to take a direct interest in the essential and distinctive characteristics of the said concept by approaching them from various angles of analysis. This makes it possible to define an ontology and its concepts through different facets that cannot be represented in the Protégé environment. We can mention, for artefacts in the field of IT, the technical axis (mechanical, electromechanical, electronic), the theoretical axis (analog, logic, or digital), the axis by program (programmed, partially programmed, non-programmed), by architecture (Von Neaumann, Harvard, parallel, quantum, ...), etc. Thus, a Von Neumann-type architecture computer is an electronic calculating machine, both arithmetic and logical, with a program, in main memory during execution. This modeling of IT heritage knowledge complements the first-level ontology by integration via some CIDOC-CRM upper-ontology classes (Fig. 3).

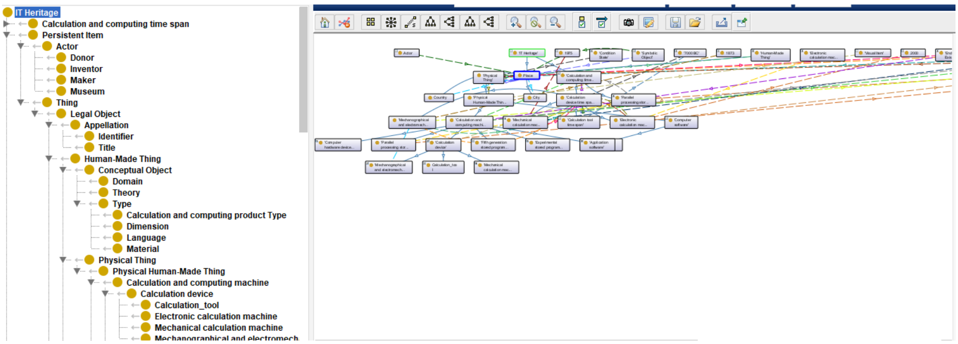


Figure 3: an overview of the ITinHeritage domain ontology.

We aligned on Protégé, the CIDOC-CRM upper-ontology to the OWLTime model (Pan and Hobbs 2005) to obtain a diachronic modelization representing the technological evolutions in the IT field. This way, the concept is located both in the global notional system and in his historical contextualization, crucial to its understanding and appropriation as explained supra. As the social context is mandatory too, our future works will focus on practices across history with the help of the “Use” class, to situate these artefacts as socially founded technologies that shape and impact our society in reciprocal ways.

Our future developments will aim to improve the diachronic representation by using Allen intervals instead of OWLTime in the TEDI environment. We are therefore overcoming the fixed-in-time aspect, which is a major bias in ontologies, by making the ITinHeritage ontology diachronic. But domain ontologies also often have the drawback of being scientific works considered as an end in themselves. However, our aim here is to demonstrate their usefulness in overcoming the difficulties of managing large masses of museum data and even beyond. Our ontology must, therefore, be linked to the data to be managed, that is to say, to the more than 25500 artefacts that make up our knowledge graph. So, we need to populate our ontology, which means assigning instances (real-world objects) to our classes (concepts). Obviously, this is impossible to be done manually in such proportions.

As a first intention, we used statistics with TF-IDF (term frequency-inverse document frequency), which measures the importance of a word to a text adjusted for the fact that some words appear more frequently in general, to automatically populate our ontology, i.e. to integrate the museum artefacts as instances of the classes (concepts) in the ontology. Here, again, lexical extraction techniques using the Sketch engine tool were necessary. It aimed to produce a lexicon for each concept in each language (FR, EN, IT, DE), using, this time, a wider corpus composed of web sources as Wikipedia and open access specialized websites or revues. This work still requires extensive human intervention with uncertain results, as TF-IDF is still a relatively old technique. And our representation of domain knowledge remains static.

We need to automate our ontology populating so that, via the synchronization our knowledge graph with the museum data (see above), we can pass on any new entries in the collections. Moreover, our ontology has to be evolutionary because IT cannot be represented in a fixed way when it is continually being created. This is why we are using Artificial Intelligence (AI) via Large Language Models (LLMs), in this case Llamma3, to automatically extend the manually constructed onto-terminology. The lexicons produced previously (we have only used the English language here) were used to guide the AI’s results. To train it, we had to provide it with descriptions representative of the domain’s terminology, corresponding to each concept. The AI’s contextualization and learning capabilities enable it to offer a classification that is refined over time. But the upstream human work required to obtain an operational AI and ontology is still very heavy in terms of lexicon construction, annotation of data corpora, and learning, and it requires the intervention of experts in the field. However, while these problems are not trivial, they are generic to the use of AI and are in the process of being improved. Our contributions here, therefore, concern the application of automatic methods for extending manually constructed ontologies in order to keep pace with technical, terminological and conceptual developments in the field. Our terminology work is being supported by the LLM to analyze textual fields in the Knowledge Graph corpus, enriched with web data sets. The aim is to be able to detect new terms relating to new objects and integrate them into our ontology, albeit still under human validation.

Finally, since the aim of all this work is to ensure that the general public has access to knowledge in the IT field, we will be making access to the Knowledge Graph possible via a web platform (www.itinheritage.com). Other portals federate museum collections (Szekely et al. 2013; De Boer et al. 2013; Doerr et al. 2010) or safeguard software (Di Cosmo 2018), but the aim of these projects is to enhance the value and heritage of artefacts, whereas our aim is to transfer knowledge. The web platform has therefore been designed with an end-user approach and offers navigation via a time map to represent the technological evolution across history (time and space representation); a visual graph based upon the domain ontology to provide input to the knowledge graph via concepts representing the domain knowledge; and a natural language querying. This man-machine interface is essential because querying the Knowledge Graph (which is in RDF) in natural language requires SPARQL (SPARQL Protocol and RDF Query Language) queries, a computer language that users in the general public cannot master. SparNatural (Clavaud and Francard 2022) is a well-known open-source tool available for querying by non-experts, which can be used to navigate our knowledge graph based on our domain ontology. This approach significantly lowers the barrier to entry for users unfamiliar with technical query languages, making the data within our Knowledge Graph accessible to a broader audience. In this way, we try to increase the possibilities for transferring knowledge of the IT field to the widest possible audience, from scientists to the general public. Firstly, by opening up and then linking data on the semantic web, and secondly by making this data intelligible to introduce specialized concepts into culture. Our future work aims to improve these ontology extension and knowledge access technologies, using a Knowledge RAG (Retrieval Augmented Generation) (Lewis et al. 2020), another technology that also combines AI and Natural Language Processing (NLP), which will enable natural language querying in a much more intuitive and efficient way, and multilingual issues to be taken into account automatically, freeing up a significant amount of human intervention in lexicon generation.

Conclusion

For Smith (2011) heritage is defined in the tension between the objects, languages and representations that form it. In this article, we have shown that this is indeed the case for Information Technologies (IT), which are gradually being elevated to the rank of heritage beyond the everyday uses they represent for us. The uses of its languages fluctuate in step with technological developments and the conceptualizations that are created by the practices around these new technologies. The aim of our work is to highlight and analyze these parallel developments. In this continuity, our future works will continue, through the linguistic approach, analyzing the diachronies and diastracies in IT heritage languages, in a greater depth, for example, by using R language tools on larger comparative corpora. We will combine and enhance too, the terminological work with Artificial Intelligence (AI), to determine its contribution to the modelling of the extra-linguistic layer, always following these developments. By digitally perpetuating IT heritage in both its tangible and intangible forms, and opening it up, we hope that our work will help the general public to gain a better understanding of this prolific and highly technical field. From a scientific point of view, our contributions aim to offer a methodology for the cultural world and beyond for: managing massive and heterogeneous data, opening and segmentizing it (Linked Open Data (LOD)); creating onto-terminologies of scientific and technical domains; and using ultimate AI tools to complement human terminology work. Finally, from an epistemological perspective, the ITinHeritage research project is developing this work to bring to light the new knowledge and representations of the world brought by IT, which are, far beyond a technoscience, the reflection of our societal mutation.

Acknowledgments

This work has been partially supported by MIAI @ Grenoble Alpes, (ANR-19-P3IA-0003).

References

Anthony, Laurence. 2021. "Programming for corpus linguistics." In A practical handbook of corpus linguistics, edited by M. Paquot, and S. T. Gries. Springer. https://doi.org/10.1007/978-3-030-46216-1_9.

Apel, Karl Otto. 1987. *L'éthique à l'âge de la science*. Presses Universitaires de Lille.

Aristotle. 1986. *Métaphysique*. Librairie Philosophique J. Vrin.

Bekiari, Chryssoula, George Bruseker, Martin Doerr, Christian-Emil Ore, Stephen Stead, and Athanasios Velios. 2021. "CIDOC CRM." International Committee for Documentation (CIDOC) of the International Council of Museums (ICOM). Version 7.1.1.

Bird, Steven, Ewan Klein, and Edward Loper. 2009. *Natural language processing with Python: Analyzing text with the Natural Language Toolkit*. O'Reilly Media.

Clavaud, Florence, et Thomas Francart. 2022. "Sparnatural, un éditeur graphique souple et intuitif pour explorer des graphes de connaissances." *Colloque Humanistica 2022*.

De Boer, Victor, Jan Wielemaker, Judith van Gent, et al. 2013. "Amsterdam museum linked open data." *Semantic Web* 4: 237-43. https://doi.org/10.3233/SW-2012-0074.

Di Cosmo, Roberto. 2018. "Software heritage: why and how we collect, preserve and share all the software source code." *IEEE/ACM 40th International Conference on Software Engineering in Society (ICSE-SEIS), Gothenburg, Sweden.*

Diels, Hermann. 1903. *Die fragmente der Vorsokratiker griechisch und deutsch.* Weidmann.

Dijkshoorn, Chris, Lizzy Jongma, Lora Aroyo, et al. 2018. "The Rijksmuseum collection as Linked Data." *Semantic web* 9 (2): 221-30. https://doi.org/10.3233/SW-170257.

Djambian, Caroline, Giada D'Ippolito, et Micaela Rossi. 2024a. "Des représentations du monde des Technologies de l'Information : un travail terminologique en patrimoines scientifiques et techniques." In *Proceedings of TOTh 2024-Terminologie & Ontologie: Théories et applications, June 2024, Le Bourget du Lac, France*, sous la direction de by Christophe Roche. Presses Universitaires Savoie Mont Blanc. https://hal.science/hal-04840125v1.

Djambian, Caroline, Micaela Rossi, Giada D'Ippolito, et al. 2024b. "New terminological approaches for new heritages and corpora: the ITinHeritage project". In *Proceedings of the 3rd international conference on Multilingual digital terminology today. Design, representation formats and management systems - MDTT 2024*, edited by Federica Vezzani, Giorgio Maria Di Nunzio, Beatriz Sánchez Cárdenas, et al. https://ceur-ws.org/Vol-3703/.

Djambian, Caroline. 2011. "Le métier : son savoir, son parler." In *Actes de la Conférence TOTh 2011-Terminologie & Ontologie: Théories et applications, May 2011, Annecy*, sous la direction de Christophe Roche. Presses Universitaires Savoie Mont Blanc. https://hal.science/hal-00805596v1.

Doerr, Martin, Stefan Gradmann, Steffen Hennicke, Antoine Isaac, Carlo Meghini, and Herbert Van de Sompel. 2010. “The Europeana Data Model (EDM).” In *Proceedings of IFLA 76.*

Drouin, Patrick. 2003. “Term extraction using non-technical corpora as a point of leverage.” *Terminology. International Journal of Theoretical and Applied Issues in Specialized Communication* 9: 99-115. https://doi.org/10.1075/term.9.1.06dro.

Felber, Helmut. 1987. *Terminology Manual.* UNESCO.

Gaudin, François. 1996. “Dire les sciences et décrire les sens : entre vulgarisation et lexicographie, le cas des dictionnaires de sciences.” *TTR: traduction, terminologie, redaction* 8: 11-27.

Gehrke, Stefanie, Eduard Frunzeanu, Pauline Charbonnier, and Marie Muffat. 2015. “Biblissima’s Prototype on Medieval Manuscript Illuminations and their Context.” In *Proceedings of International Workshop of Semantic Web for Scientific Heritage at the 12th ESWC 2015 Conference, Portorož, Slovenia, June 1st, 2015*, edited by Arnaud Zucker, Isabelle Draelants, Catherine Faron-Zucker and Alexandre Monnin, CEUR Workshop Proceedings, 43-48, CEUR-WS.org. http://ceur-ws.org/Vol-1364/paper5.pdf.

ISO. 2025. *ISO 5078:2025 Management of terminology resources — Terminology extraction*.

ISO. 2019. *ISO 1087-1. Travaux terminologique – Vocabulaire – Partie 1 : Théorie et application*.

ISO. 2009. *ISO 704 NF ISO 704. Travail terminologique – Principes et méthodes*.

Jacob, Christian. 2001. “Rassembler la mémoire.” *Diogène* 4: 53-76.

Juanals, Brigitte, et Jean-Luc Minel. 2016. “La construction d’un espace patrimonial partagé dans le Web de données ouvert.” *Communication* 34 (1). https://doi.org/10.4000/communication.6650.

Kant, Immanuel. 1997. *Critique de la raison pure.* Aubier.

Kilgarriff, Adam, Vít Baisa, Jan Bušta, et al. 2014. “The Sketch Engine: Ten years on.” *Lexicography* 1: 7-36.

Kripke, Saul. 1982. *La logique des noms propres*. Les Editions de Minuit.

Lewis, Patrick, Ethan Perez, Aleksandra Piktus, et al. 2020. “Retrieval-Augmented Generation for Knowledge-Intensive NLP Tasks.” *Advances in Neural Information Processing Systems* 3: 9459-74. https://doi.org/10.48550/arXiv.2005.11401.

Lafont, Robert. 1993. *Le dire et le faire*. Praxiling.

Lefeuvre, Luce, et Anne Condamines. 2017. “MAR-REL : une base de marqueurs de relations conceptuelles pour la détection de Contextes Riches en Connaissances (MAR-REL : a conceptual relation markers database for Knowledge-Rich Contexts extraction).” In *Actes des 24ème Conférence sur le Traitement Automatique des Langues Naturelles. Volume 2 - Articles courts, Orléans, France. ATALA*.

Lerat, Pierre. 1987. “Le traitement des emprunts en terminographie et en néographie”. *Cahiers de lexicologie* 50: 137-44.

Lerat, Pierre. 1988. “Terminologie et sémantique descriptive.” *La banque des mots* 36.

Pan, Feng, and Jerry R. Hobbs. 2005. “Temporal aggregates in owl-time.” In *Proceedings of the Eighteenth International Florida Artificial Intelligence Research Society Conference (FLAIRS 2005).*

Porphyre. 1984. *Isagoge*. Librairie Philosophique J. Vrin.

Putnam, Hilary. 1984. *Raison, vérité et histoire*. Editions de Minuit.

Putnam, Hilary. 1990. *Représentation et réalité*. Gallimard.

Rey, Alain. 1990. “Lexico-logiques, discours, lexiques et terminologies ‘philosophiques’.” *Encyclopédie philosophique universelle. T. II. Les notions.* Presses Universitaires de France.

Roche, Christophe. 2007. “Le terme et le concept : fondements d’une ontoterminologie.” In *Actes de la Conférence TOTh 2007-Terminologie & Ontologie: Théories et applications*, *Annecy 1er juin 2007*, sous la direction de Christophe Roche. Presses Universitaires Savoie Mont Blanc.

Roche, Christophe, and Maria Papadopoulou. 2019. “Mind the Gap: Ontology Authoring for Humanists.” In *Proceedings of JOWO: The Joint Ontology Workshops*, edited by A. Barton, S. Seppälä and D. Porello. https://ceur-ws.org/Vol-2518/.

Roche, Christophe. 2005. “Terminologie et ontologie.” *Langages* 1: 48-62.

Rossi, Micaela, Caroline Djamban, Cécile Frérot, Giada D’Ippolito, Emrick Poncet. En cours de publication. “Rôle et apport des verbes pour la construction et la modélisation de connaissances spécialisées du patrimoine muséal.” In *Actes du Colloque Le statut du verbe terminologique entre théories et pratiques*.

Russell, Bertrand. 1956. *Logic and Knowledge*. Allen and Unwin.

Simon, Agnès, Sébastien Peyrard, Vincent Michel, and Adrien Di Mascio. 2014. “We grew up together: data.bnf.fr from the BnF and Logilab perspectives.” In *Proceedings of IFLA 2014*.

Smith, Laurajane. 2011. “Heritage and its Intangibility.” In *De l’immatérialité du patrimoine culturel*, edited by A. Skounti and Ouidad Tebbaa. Bureau régional de l’Unesco de Rabat.

Szekely, Pedro, Craig A. Knoblock, Fengyu Yang, et al. 2013. “Connecting the Smithsonian American Art Museum to the Linked Data Cloud.” In *The Semantic Web: Semantics and Big Data. ESWC 2013,* edited by P. Cimiano, O. Corcho, V. Presutti, L. Hollink, S. Rudolph. *Lecture Notes in Computer Science* 7882. Springer, Berlin, Heidelberg. https://doi.org/10.1007/978-3-642-38288-8_40.

UNESCO. 2009. *Guide to measuring Information and Communication Technologies (ICT) in education*. UNESCO. https://uis.unesco.org/sites/default/files/documents/guide-to-measuring-information-and-communication-technologies-ict-in-education-en_0.pdf.

Wittgenstein, Ludwig. 1961. *Investigations philosophiques*. Gallimard.

AIDAinformazioni

Rivista semestrale di Scienze dell'Informazione

Anno 42

N. 3-4 – luglio-dicembre 2024

Contributi

Rubriche

mundaneum

In copertina
Disegno di Paul Otlet, Collections Mundaneum, centre d'Archives, Mons (Belgique).

ISBN 979-12-5965-549-3 ISSN 1121-0095

9 791259 655493 9 770112 100950