Article

em questão
E-ISSN 1808-5245

Knowledge organization for resilience in times of crisis: challenges and opportunities[1]

Birger Hjørland [I]

https://orcid.org/0000-0001-9535-1732

[I] University of Copenhagen, Copenhagen, Denmark

Abstract: This paper considers different kinds of information systems, knowledge organization systems and information infrastructures in relation to problems in times of crisis, such as the COVID-19 pandemics. Considering such issues quickly turns into more fundamental issues in knowledge organization. Among the systems discussed are general infrastructures for the general public, such as Wikipedia, Google, and ChatGPT, domain specific structures for the general public such as health-related websites, general structures for the specialists, such as the citation databases Scopus, Dimensions, Google Scholar and Web of Science, and domain specific structures for the specialists such as MEDLINE and FindZebra. All these systems are considered exemplified mostly from the COVID-19 perspective. The paper focus on the common problem for all kinds of knowledge organization systems, including the issues about "document retrieval versus "information retrieval," "natural language" vs "controlled vocabulary", "best match" vs "exact match" and issues related to the philosophy of science. It is concluded that in times of crisis, both specialized new systems and updating of existing systems are needed. Most important for knowledge organization as a field of research and practice is the focus on the common theoretical foundation of all kinds of knowledge organization systems because this is what can keep us together.

Keywords: Knowledge Organization Systems; information infrastructures; resilience; theoretical issues

1 Introduction: The complex field of knowledge organization

Knowledge organization (KO) is about making, evaluating and managing different kinds of knowledge organization systems (KOS) or information infrastructures.[2] There are many kinds of KOS/infrastructures, they may be

classified in two dimensions: (1) General structures encompassing all fields of knowledge versus specialized structures encompassing some specific fields; (2) Information structures aimed at the general public versus structures aimed at a specific audience. Chart 1 provides a few examples:

Chart 1 - An overall classification of KOS/information infrastructures

	General structures	**Domain specific structures**
Information structures for the general public	(1) E.g., Wikipedia, Google, ChatGPT, Library of Congress Subject Headings	(2) E.g., CDC COVID-19 site
Information structures for specialists	(3) E.g., Scopus, Dimensions, Google Scholar and Web of Science	(4) E.g., MEDLINE and FindZebra.

Source: Made by the author.

The field of KO covers a very large terrain: General KOS and KOS in all domains of knowledge, both aiming at the general public and at the specialists. To this all the different kinds of technologies and approaches in use or to be discovered, and the great number of interdisciplinary perspectives, including statistical, linguistic, philosophical, and sociological.

As if not all this were enough, societal development and crises such as the Corona-19 pandemic contribute new challenges. Donald Trump's USA has created the wave of "fake news" and distrust in science and other authorities, while the Corona-pandemic, which KOS and information infrastructure are meant to combat, created new challenges for KO, such as the phenomenon known as "Infodemics."

WHO (2023) defined:

> An infodemic is too much information including false or misleading information in digital and physical environments during a disease outbreak. It causes confusion and risk-taking behaviours that can harm health. It also leads to mistrust in health authorities and undermines the public health response. An infodemic can intensify or lengthen outbreaks when people are unsure about what they need to do to protect their health and the health of people around them. With growing digitization – an expansion of social media and internet use – information can spread more rapidly. This can help to more quickly fill information voids but can also amplify harmful messages.

Segault (2022) found:

> Like every large-scale crisis, the COVID-19 pandemic is also an information crisis: new questions, new information, new uncertainties, new rumours and new lies have been appearing all at once and have been travelling at a fast pace in a globalised information ecosystem (Segault, 2022, p. 37).

This complex field of KO gives rise to a wide range of specialized investigations. Some of them are about strategies to cope with issues like vaccine skepticism – e.g., Gallagher and Lawrence (2020), Claggett *et al*. (2022). We shall not here go into such strategies, just focus on infrastructures providing the best available knowledge. In this presentation, some examples are given. However, underlying all such specialized investigations are some core theoretical and philosophical problems, which, it is argued, cannot be avoided. It is only these underlying problems, which provide means to see what is common for all KOS, and what may provide a basis for KO to be a coherent field rather than a scattering of a long range of separate fields.

2 Resilience in research in information science

The term "resilience" has been used in research in information science, and in other fields. De Fremery, Buckland and Le Deuff (2022) found:

> An information-resilient society requires the development of a resilient information science. Three different requirements for resilient information science are identified and a constructive contribution presented for each. First the scope and available resources need to be fully identified. […] Second, clarity in concepts and definitions is required. […] Third, concepts and techniques need to be articulated as new technologies emerge. (De Fremery; Buckland and Le Deuff, 2022, p. 559)

Some other suggestions are in line with this view, for example, Dickey *et al*. (2022), who described recent standards revision, development, and initiatives by NISO, ISO and W3C focusing on accessibility and information resilience. Papajorgji and Pardalos (2014, p. 191) emphasized that because implementation technologies change constantly, it is highly desirable that the models we develop are expressed in an abstract and logical manner resilient to changes, recommending the Unified Modeling Language (UML).

Other suggestions emphasized the broader social issues. Brannon *et al*. (2022) found that part of contributing to an information-resilient society is being aware of the many ways that society censors or limits access to ideas and knowledge, and Reynolds *et al*. (2022) wrote:

> A focus on the social and ecological dimensions of core traits of resilient systems – diversity, redundancy, modularity, connectivity, regenerative ability, and equitability – yields a framework to guide the inventory and analysis of UGI toward greater resilience. (Reynolds *et al*., 2022, p. 237)

Mazzocchi (2021) is an article about drawing lessons from the COVID-19 pandemic. He pointed out that the pandemic has widely spotlighted disagreements among scientific experts because they had to deal with a novel, highly complex problem. New data sets have been continuously collected and analyzed, circulating even before being peer-reviewed through pre-print services. He found that an important learning has been that "science and epistemic humility should go together". Experts should recognize the uncertainty that might accompany their claims, adjusting their epistemic conduct and communication accordingly. Mazzocchi also argued for open debates in the media about uncertainty, against other researchers who believe that the price to be paid for such disputes will be a drop of public trust in scientists, as a few surveys seem to show.

In addition to these views from the literature, it may be suggested that resilience of information infrastructures needs to consider the public-private cooperation. Traditionally public libraries where the central infrastructure, but today private services like Wikipedia, Google etc. dominate.

3 Examples of research on information infrastructures

This section presents different information structures with emphasis on exemplifying their treatment of COVID-19 information. The overall subsections are 3.1: General structures for the general public; 3.2: Domain specific structures for the general public; 3.3: General structures for specialists and 3.4: Domain specific structures for specialists.

3.1 General structures for the general public

This subsection discusses the following information infrastructures: 3.1.1: Wikipedia: 3.1.2: Google and 3.1.3: ChatGPT.

3.1.1 Wikipedia

There is much research about Wikipedia, no other encyclopedia has ever been examined so closely and in so many different ways, and much of this research is referred to by Wikipedia itself, making Wikipedia an important source of information about e.g., the reliability of its information and about Wikipedia's coverage of the COVID-19 pandemic. Overall, Wikipedia is highly self-reflexive in its procedures, norms and discussions about itself. Wikipedia (2024b) wrote:

> The COVID-19 pandemic was covered in Wikipedia extensively, in real-time, and across many languages. This coverage extends to many detailed articles about various aspects of the topic itself, as well as many existing articles being amended to take account of the pandemic's effect on them (Wikipedia, 2024b).

This seems to be a correct description. This author found on November 19, 2023, that there were 123,909 articles in the English Wikipedia containing the term "COVID-19". Deltell and Claes (2021) found that Wikipedians have reacted quickly to the pandemic and generated great collaborative work, with articles argued and constructed from a scientific and health perspective. Benjakob *et al*. (2022) found that "coronavirus-related articles referenced trusted media outlets and high-quality academic sources" and "it successfully fended off COVID-19 disinformation".

Segault (2022) wrote: "As the encyclopaedia evolved and the media crisis unfolded, the press coverage of the encyclopaedia went through different phases, from scepticism to acceptance, and even trust" (Benjakob; Harrison, 2019). At a time when traditional media outlets proved vulnerable to misinformation (due to time pressure and cost cuts), sensationalism (through audience races) and manipulation (by political discourses or economical agreements), the specificities of Wikimedia were recognised as strengths".

In the overwhelming amount of research on Wikipedia, an important focus must be on how to contribute to the possible improvement of Wikipedia or its use.

Given what I formerly wrote, there seems not to be any obvious or easy problem to solve. For now, I have two suggestions:

a) A classified index to the contents of Wikipedia (in our example: A classification scheme for the many entries about COVID-19 and related topics.)

Although Wikipedia already has many category-systems and internal links, it is often difficult to get a clear overview of a topic – the category "COVID-19" (Wikipedia, 2024a), for example, does not provide a good overview of entries. Principles on how to construe this index can be found in Facet analytic methods (at least partially).

b) Research on uncovering bias in Wikipedia by examine the representation of different views on a topic (here: COVID-19) and subject qualifications as well as other attributes of editors and writers.

Given what is already said, this is not an easy task. Wikipedia does not, like traditional high-quality encyclopedias provide signed, but anonymous articles (although pseudonym and IP-address are given for each change made just as many Wikipedians have a personal page providing some information). This means that users cannot identify the authors and check their qualifications and publication record. The best way seems to be to examine the coverage of different views/perspectives in any domain.

3.1.2 Google

Google treats topics related to health, safety, financial stability etc. as a category YMYL (Your Money or Your Life) for which expertise, authority and trustworthiness is especially important for the 380.000 raters[3], which, according to Sundin *et al*. (2023) are employed by Google. Google use these raters to ensure that what Google considers to be lower-quality health and medical articles become less visible to users. In Google (2023) is written:

> Respect user choice: Users who express an intent to explore content that is not illegal or prohibited by our policies should be able to find it, even if all available indicators suggest it is of relatively low quality. We set a higher bar for information quality where users have not clearly expressed what they are looking for.

In other words, Google has quality criteria different to both kinds of content (YMYL vs. non YMYL) and in relation to how users express what they are looking for. Sundin *et al*. (2023) performed a study of how drugs for treating COVID-19 could be found in Sweden. The study distinguished drugs recognized by health authorities versus drugs that were "gray" (i.e., not recognized). None of the gray drugs were available from official pharmacies but might be bought from "grey-zone pharmacies".

Among the grey drugs, the one named "ivermectin" attracted the highest interest among Google searchers. A search for this name provided a result in which all but one of the first ten links lead to what is commonly regarded as trustworthy websites providing information from credible sources. *Here, Google's YMYL principle seems to work satisfactorily*.

However, when searching "ivermectin köpa [buy]", the first ten links, with one exception, are links to sites outside the officially sanctioned universe of health and medical websites. The rest of the links are to various pharmacies in the grey zones from which it is possible to order ivermectin as a cure for COVID-19. *Here, another of Google's principle is at play: "Respect user choice"*.

Sundin *et al*. (2023) also documented clear signs of "malicious redirections", which are bits of code, not visible to the ordinary user, that are inserted into the core files of a website. They are designed to divert website visitors to a specified, unrelated site that often contains adverts, pornography, etc. Moreover, the article shows how the grey-zone pharmacies draw on malicious behavior to bypass the algorithmic intention behind prioritizing quality of YMYL pages, to become visible.

This study is interesting in relation to the theme of this conference, demonstrating that a system like Google is not just driven by algorithms, but also of *human classifications based on ethical principles*. In addition, it highlights the problem that outsiders may manipulate search engines to counteract their ethical and other principles.

Two general issues in relation to search engines should be mentioned:

a) A general system like Google prioritizes the most dominant views, whereas searches for seldom views may be less well served. A system called FindZebra documented that building specialized search engines rather than general ones provides better searches for rare diseases – Dragusin *et al*. (2013a), Dragusin *et al*. (2013b), Liévin *et al.* (2023);
b) Modern IR is dominated by "best match" rather than "exact-match retrieval" This, however, although far more user friendly, deprives the searcher of transparency and control over the search process and deprive informational professionals the well-established strategies to increasing recall and precision. Perhaps "classical databases" like MEDLINE should not be considered obsolete (although just fulfilling some special needs). More on this in Section 4.

3.1.3 ChatGPT

ChatGPT (Chat Generative Pretrained Transformer) is a system based on artificial intelligence (AI), more specifically the technology known as large language models (LLM). It has become extremely hyped and expected to revolutionize many processes, including IR, and KO (e.g., by automatically producing ontologies and other KOS). Before discussing its theoretical problems, let us consider a specific example on how it works:

Chart 2 - A dialog with ChatGPT-4

Hjørland: "Do you know of empirical comparisons of FindZebra versus Google?"
ChatGPT 4: "As of my last update in January 2022, I don't have specific empirical comparisons between FindZebra and Google in terms of their performance for medical information retrieval. However, I can offer some insights into potential differences between the two."

ChatGPT 4 continued with a range of common sense and hypothetical differences between FindZebra and Google, such as: "FindZebra may offer higher precision and recall for rare disease queries compared to Google," which are not properly documented answers. Its final conclusion was: "Empirical comparisons between FindZebra and Google would involve evaluating metrics such as precision, recall, user satisfaction, and efficiency in retrieving relevant medical information, particularly for rare diseases. Conducting such studies would require access to usage data from both platforms, user feedback, and defined criteria for evaluating search performance."

Hjørland "But the articles "*FindZebra: A search engine for rare diseases*" and "*Specialized tools are needed when searching the web for rare disease diagnoses*" from 2013 contains empirical information about the advantages of FindZebra compared to Google. These articles are open on the web, but you did not use their findings?"

ChatGPT 4: "I appreciate your patience. Let me address your question with a more focused response, drawing upon the empirical information from the articles you mentioned.

- The articles […] indeed provide empirical insights into the advantages of FindZebra compared to Google for rare disease searches. Here are some key findings from these articles: 1. Precision and Recall:
- FindZebra demonstrated higher precision and recall for rare disease searches compared to Google.
- Google's search results often contained a significant amount of irrelevant or non-authoritative information for rare diseases, leading to lower precision and recall. (Continuing with a more detailed list of findings from the articles)

Source: Research data.

This example shows that ChatGPT did not find the relevant empirical articles when first needed (even claiming it did not have access to such studies).

When articles containing relevant empirical knowledge was mentioned to ChatGPT, it was seemingly able to find the articles and summarize their findings in a few seconds, but without citing accurate findings (no quantitative data were provided). Overall, ChatGPT-4 did not provide satisfactory answers on an academic level.

A main problem seems to be ChatGPT's lack of relating its claims with the sources from which these claims are obtained.

I asked it to provide a list of 10 bibliographical references authored by Birger Hjørland about domain analysis. First it provided a description of three papers, without satisfactory bibliographical data. Asking again, it provided a list of 10 bibliographical references. I asked to have them formatted according to APA standard, and it correctly did so immediately.

However, of the 10 refs, most were not about domain analysis and just two were by Hjørland. One was pure "fiction" (there is no such thing as "Hjørland, B. (2016). Knowledge organization: A new science? Knowledge Organization, 43(8), 641-655."[4].

The problem of relating claims to the sources was also mentioned by Briganti (2024), who wrote:

> It [ChatGPT] is trained on a diverse range of internet text, but it does not know specifics about which documents were part of its training set. (Briganti, 2024, p.1565)

This seems to be a core problem: ChatGPT answers are not connected to a specific document (or source) but is based on statistical relations among words across documents. Remark that traditional IR systems in reality are document-retrieval systems (despite the name "information retrieval systems"), but that ChatGPT is a system which provides information rather than documents (unless you specifically ask for documents, and in that case, they are often "bibliographical ghosts").

I consider a system useless for IR, if you do not know whether you can trust its answers or not, as in the last example. The problem with

incorrect or nonsensical answers ("hallucinations") in ChatGPT is well known and was admitted by its producer, OpenAI (2022), who wrote:

> ChatGPT sometimes writes plausible-sounding but incorrect or nonsensical answers. Fixing this issue is challenging, as: (1) during RL training [Reinforcement Learning], there's currently no source of truth; (2) training the model to be more cautious causes it to decline questions that it can answer correctly; and (3) supervised training misleads the model because the ideal answer depends on what the model knows, rather than what the human demonstrator knows.

ChatGPT-4 wrote in 2024: "The challenges described in the statement you provided are indeed representative of some of the difficulties faced in training AI models like ChatGPT. As of my last update in April 2023, the issues of generating plausible-sounding but incorrect answers, the lack of a "source of truth" during certain types of training, and the balance between cautiousness and helpfulness were still areas of active research and development. OpenAI has been continuously working on improving AI models to reduce the frequency of such issues". (Followed by a list of initiatives to cope with this problem).

It must be said that ChatGPT-4 has higher accuracy compared to the free ChatGPT-3.5[5]. This could perhaps indicate that the underlying principles are sound, that it just need to be further fine-tuned. This is an important question, which to me seems related to the issue about the relation between claims and sources, and therefore to the question of whether document or information should be retrieved. I'll return to this issue in Section 4.

Another important question is: Does ChatGPT understand texts, concepts etc.? Or does it just rely on examples in the database? The algorithms as well as the training documents are secrets, but Yousefzadeh and Cao (2023) found:

a) "It is hard to find scientific evidence suggesting that GPT-4 has acquired an understanding of even basic mathematical concepts";

b) "Our finding suggests that GPT-4's ability is to reproduce, rephrase, and polish the mathematical proofs that it has seen before, and not in grasping mathematical concepts";
c) and "predicting the next word in a sentence may be a misguided approach, a recipe that often leads to excessive extrapolation and eventual failures".

Let us finish ChatGPT by considering its relation to knowledge organization. As far as I know, systems like ChatGPT have not improved on core principles or problems related to IR and KO, but they just rely on a very large text corpus and the ability to provide coherent answers in well-written languages. Therefore, my view is that such AI systems have so far not challenged research in IR and KO. As Neuhaus (2024) concluded:

> [W]e need to be aware that LLMs do not implicitly hide an ontology within their billions of parameters, which we just need to tap into. The training data of any LLM likely contains a variety of divergent views on any given subject matter and LLMs are not designed to ensure logical consistency. Therefore, we should not expect the output of an LLM to reflect a logically consistent and ontologically sound view of any given domain" (Neuhaus, 2024, p.10)

3.2 Domain specific structures for the general public

These are, for example, the many health-oriented homepages meant for the non-specialists. One example is *Centers for Disease Control and Prevention* (CDC), e.g., on COVID-19 (CDC, 2025).

In general, such structures probably serve their purpose well. A main problem is that there are many, that there is no overall coordination, and that it therefore may be a bit arbitrary what information is disseminated to the users. In addition, commercial interests may be at play at many sites, and users may find it difficult to choose the best.
I will not say more about this category.

3.3 General structures for specialists

Important examples are the citation databases. After date launched, the most important ones are:

a) *Web of Science* (WoS 1997) (Going back to *Science Citation Index*, 1964);
b) *Scopus* (2004-);
c) *Google Scholar* (2004);
d) *Dimensions* AI (2018-).

Science Citation Index challenged traditional disciplinary databases by not applying human classification and indexing. Scopus became bigger than WoS. *Google Scholar* went a step further and used automatic collection and organization of its data, and has a better coverage than, e.g., MEDLINE. *Dimensions AI* have added many interesting futures, including new kinds of documents covered and subject indexing at the article level.

Systems like *Google Scholar* is today challenging systems like *PubMed/Medline*, but to my knowledge there is no evidence they are superior. *Cochrane Handbook for Systematic Reviews of Interventions* (Lefebvre *et al*., 2023) still focus on traditional databases: "Ensure that CENTRAL, MEDLINE and Embase […] have been searched," whereas computer scientists often are sceptical of traditional databases.

For the future of KO, this is an important question: What are the benefits and drawback of different kinds of human based vs algorithmic solutions. We should remember that just as IT can be improved, so can systems like PubMed. I believe that better theoretical foundation of KO is essential for further developments. We shall return to the nature of such theoretical issues.

3.4 Domain specific structures for specialists

Zeng *et al*. (2020) provided a comprehensive overview of KOS for health communication during the COVID-19 pandemic. The article covered both the major health KOS (e.g., International Classifications of Diseases, ICD-10), five ontologies for COVID-19 research, as well as other kinds of systems and standards (e.g., knowledge graphs), and described the purposes and activities of the many KOS. The article found:

> Ultimately, the networking environment in which we live also means that we are in a big-data world. The use cases of KOSs presented here have been examined in context of their abilities to solve two major problems occurred as a result of this big-data world: first, dealing with information overload; second, resolving semantic conflicts. The fundamental approaches of KOS, including eliminating ambiguity, managing semantic relationships, and modeling ontological classes and their properties, enable effective and efficient management of information and knowledge. (Zeng et al., 2020, p. 165)

Zeng *et al*. did not consider the issues in producing and evaluating the many KOS, including the benefits and drawback of different approaches, but focused on their content and functions. Here we shall briefly look at some issues.

PubMed with MEDLINE is, as we saw, a dominant medical bibliographical database. Of special interest to the KO community is MeSH (Medical Subject Headings) and the indexing of the medical articles. How does it perform relative to other systems based on different approaches?

Bramer *et al*. (2016) found that *Google Scholar* was more complete than *MEDLINE*, while *MEDLINE* had better recall and precision (i.e., found more and had less noise). The study concluded:

> Despite its vast coverage of the scholarly literature, Google Scholar is not sufficient to be used on its own as a single database to support SR [systematic review] searching. The reason for this is not low precision in GS searching, which is comparable to traditional databases. More problematic is GS' low recall capabilities which are related to the viewable 1000 search results only policy of the search engine [...] (Bramer et al., 2016, p. 6)

This result should not be considered the final answer, as innovations in databases are continuing. What is relatively unexplored is the quality of *MeSH* and Indexing, which tends just to be taken for granted (*MEDLINE* require that indexers possess medical qualifications). We should be open for the possibility of improving the indexing and the principles on which it is based. An older study (Adams *et al*., 1994) found that adequate methodological descriptions were often missing in MEDLINE indexing. Névéol *et al*. (2010) is a study that, according to Lardera and Hjørland (2021), indicates that MEDLINE indexing may be too mechanical and too weak in the conceptualization of the articles being indexed. This is but a research hypothesis: That an alternative theoretical approach might improve retrievability of MEDLINE, but it seems an important hypothesis in times where classical databases increasingly are being challenged by systems such as Google Scholar.

MEDLINE is a specific structure compared to Google, but many infra-structures are much more specialized. This is, for example, the case with the formerly mentioned FindZebra, which specializes in rare diseases and is able to find information better than both MEDLINE and Google, although it is also drawing on MEDLINE. The theoretical most interesting questions are here:

a) the use "exact match techniques" (as MEDLINE and "classical databases") vs. the use of "best match techniques" (as in Google and FindZebra);
b) exclusively focusing on match between query and document representations (as FindZebra) versus combining such a match with other parameters, such as page popularity (e.g., using *PageRank*) and personalization (as Google does).

Here, there seems to be a conflict between "user friendliness" vs. recall. As mentioned, ISKO member Marcia Zeng has done research in this area, including research on the *International Classification of*

Diseases (ICD) (Hong; Zeng, 2022) and *Challenges, opportunities, and approaches in a health KOS vocabulary's revisions* (Zeng; Hong, 2022). At this place, attention should be directed against the problems and principles that are common for all KOS, towards which we now turn.

4 Theoretical issues underlying all KOS

In this section, just a few of my working hypotheses, which challenge mainstream views in information science are presented: 4.1 Document retrieval or "information retrieval"; 4.2 "Natural language" vs controlled vocabulary (CV); 4.3 Best match vs exact match and 4.4 Issues related to the philosophy of science.

4.1 Document retrieval or "information retrieval" ?

We saw previously that ChatGPT sometimes "hallucinates" and provides wrong answers, and that this problem in particular is related to missing documentation for its knowledge claims. Traditional IR is not about directly informing users, but about retrieving documents, which it is assumed provide answer to the "information need" of the user. Spang-Hanssen (2001) wrote:

> Information about some physical property of a material is actually incomplete without information about the precision of the data and about the conditions under which these data were obtained. Moreover, various investigations of a property have often led to different results that cannot be compared and evaluated apart from information about their background. An empirical fact always has a history and a perhaps not too certain future. This history and future can be known only through information from particular documents, i.e. by document retrieval.
> The so-called fact retrieval centers seem to me to be just information centers that keep their information sources – i.e. their documents – exclusively to themselves. (Spang-Hanssen, 2001, p. 127)

The same can be said of those using the term "information" in the same sense, and, as said, about systems like ChatGPT. There are no "sources of truth," on which such systems can verify their answers, just relative authoritative sources, which may be (or perhaps have already been) corrected by other sources. It may be more "user friendly" to provide concrete answers, but this comes at the cost of reliability and authority of the information. White (2018, p. 3927) argued likewise,

writing: "When IS [information science] is defined as the study of literature-based answering, much else falls into place".

In academic work, a core competency is to evaluate the methodology and the assumptions on which knowledge claims are made. This can only be done by document retrieval. We may use "indicators" (e.g., journal impact factors, JIFs) as a shortcut for valid knowledge, but such indicators are no guarantee for documents containing truth. In the end, experts have to evaluate knowledge claims based on the examination of documents. They may fail, or disagree, but nonetheless, this is the closest we can come to trustworthy knowledge claims, and such claims are examined and potentially revised by future documents. This is the reason *public* access to documents is essential to science. This an extremely important point.

4.2 "Natural language" versus controlled vocabulary

Controlled vocabularies (CVs) are considered important in our field of KO, but often not considered so by many computer scientists (I consider CVs as not just verbal KOSs, but also classification systems etc.) Salton (1996), for example, wrote:

> [A]cting as if we were stuck in the nineteenth century with controlled vocabularies, thesaurus control, and all the attendant miseries, will surely not contribute to a proper understanding and appreciation of the modern information science field. (Salton, 1996, p. 333)

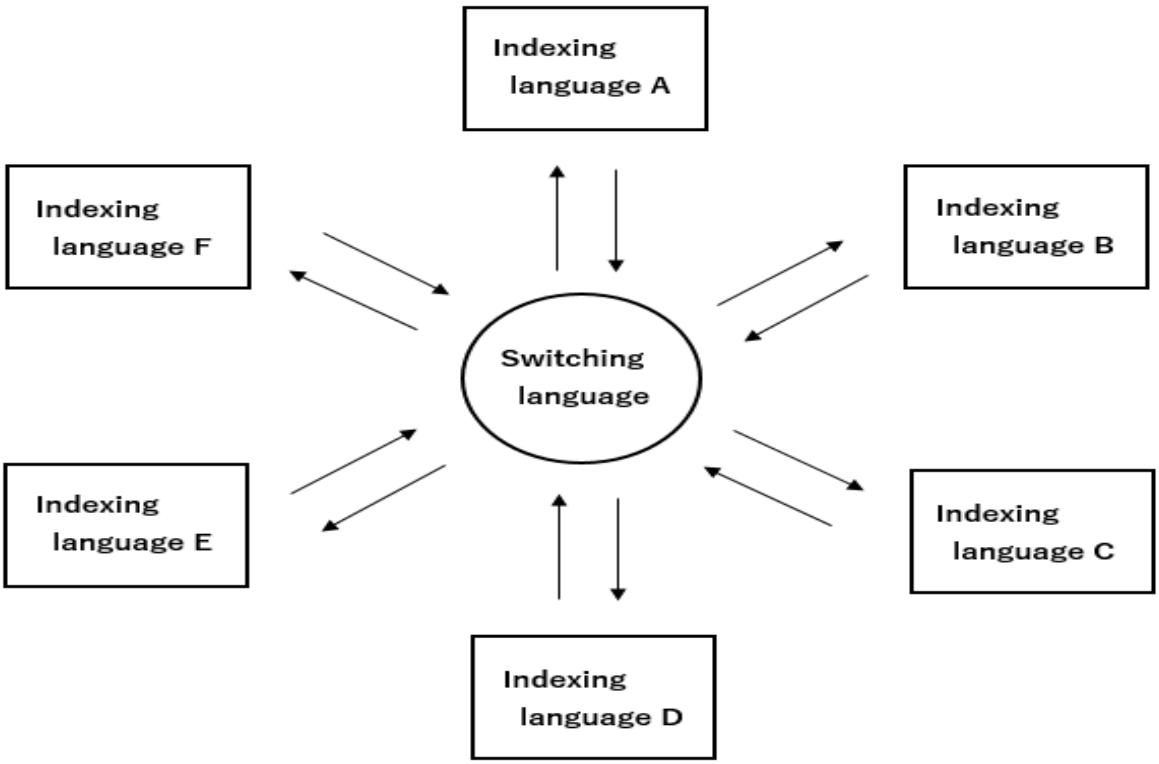
There are also important reservations about CVs from within KO. Maniez (1997), for example, wrote:

> Compatibility is the paradise lost of information scientists, the dream of a universal communication between information languages [=CVs]. Paradoxically the information languages increase the difficulties of cooperation between the different information databases. (Maniez, 1997, p. 213)

How can this be the case? One cause can be that different CVs choose arbitrary standards and thus provide problems for compatibility.

At a deeper level, the problem is associated with different interpretations of meaning. There seems in information science to be a rather naïve view about the identifications of meanings and about translating one CV to other CVs. For example, the *Broad System of Ordering* (BSO) was based on the idea that it is possible to translate different "indexing languages" (=CVs) to each other, possibly by using a "switching language":

Figure 1 – The idea of a switching language



Source: Horsnell (1975, p. 59) and Kawamura (2023, p. 9)

This idea was, however, criticized by Maniez (1997, p. 220) and does not correspond well with Kuhn's (1962) concept "incommensurability" (Oberheim; Hoyningen-Huene, 2018). In addition, claims about the translatability of CVs are seldom examined by examples from the real world. In biology, for example, a classification of organism according to structural similarities is not compatible with a classification according to common origins (genealogical classification). CVs based on

such different principles cannot be translated to each other. We have to know about this and make informed choices on this basis.

A basic theoretical issue is that any CV is based on choices about defining and relating concepts, they are not just mirroring an objective reality. Therefore, CVs are not neutral representations, but "theory laden" conceptualizations, and points of view on the world, and human choices, as biological classification demonstrates. The idea that words often reflect specific world views, meanings, values on their objects is a view formulated by Mikhail Bakhtin (1981), for which he introduced the word "heteroglossia". The "natural language" may consist of many documents, classifying, e.g., organisms, in different ways (including structural and genealogical ways), i.e., consisting of different "voices". – Recall that Neuhaus (2024, p. 10) wrote: "any LLM likely contains a variety of divergent views on any given subject".

The idea of "heteroglossia" is therefore relevant for understanding scientific languages, as demonstrated by Kuhn's (1962) theory of scientific paradigms. The implication of this view of language for information science is that we have to identify different theories and paradigms in the literature to clarify the concepts that we are organizing. Linguists, information scientists and other tend to think of language as one system, rather than as a mixture of different "voices". This is a problem for IR techniques based on natural languages, and it is also a problem when this view informs the construction of CVs. Rather than saying that a word has "a meaning" or some definite relations to other words, we should say that meanings and relations are attributed to words/signs from certain interests and theoretical perspectives.

We need KOS based on explicit theoretical criteria, not just as CVs for indexing; they will be useful for IR also when used for searching "natural language" because they help us understand and make distinctions in the domain in which we are navigating.

4.3 Best match vs exact match

As formerly said, classical bibliographical databases (such as MEDLINE) provide well defined sets of retrieved references ("exact match"), whereas modern search engines are based on "best match" statistical techniques which rank retrieved documents in order of "relevance". Only exact match provides the searcher full control over the search process.

Best match is by far the most popular, but this popularity seems to come at the potential cost of losing relevant documents. It should also be said that information professional's core competency used to be search strategies for increasing "recall" and "precision".

For example, if searching a word such as "thesaurus" provides too few hits, the truncated form "thesaur*" may increase recall (truncation is one among many "recall devices"). Searching a phrase (such as "information science" rather than just "information") may increase precision (phrase searching being one among many "precision devices"). However, as Robertson (2000) argued:

> The term precision device itself was coined in the former context [set-based retrieval]: whether it is a valid concept for the latter [ranked-based retrieval] is not obvious. To put it even more directly [. . .] Precision devices aren't what they used to be! (Robertson, 2000, p. 5)

This also means, that search strategies of the information professional "aren't what they used to be!"

I find that searching traditional bibliographical databases is closer to the theory of KO, because it is about utilizing meaning differences in titles, descriptors, full-text, etc.). It is also about considering contexts, because meanings are related to contexts, and the more a single text is merged into other contexts, the less information we have about the meanings.

Best match techniques are often used on full-text databases, whereas exact match is often used in bibliographical databases without full-text search capabilities. Best match vs exact match effectiveness

should not be confused with the relative strengths of full-text searching vs. searching databases without full-text.

4.4 Issues related to the philosophy of science

KO is influenced by many fields, including computer science, cognitive science, linguistics, and terminology studies.

However, philosophy of science has a particular important role to play, which, unfortunately, is not yet well recognized. Philosophy is about the foundations of any field of study, including computer science and linguistics. Import of knowledge from other discipline is therefore not just about some facts or established principles from these but must be based on a stance on the fields' philosophical problems (their "paradigms").

Take linguistics, for example. Spang-Hanssen (1974) wrote:

> I therefore agree with Gerard Salton in his pessimistic assessment of the relevance of the newest linguistics relevance for information retrieval. However, Salton seems to identify linguistics with newer American linguistics and thereby overlook knowledge already obtained before transformational grammar [Chomsky's approach] or at the same time in other countries, including Scandinavia. (Spang-Hanssen, 1974, p. 17; our translation)

Here is a clear expression of the view that one philosophical position in linguistics may be unfruitful for IR and KO, while others may be fruitful. For the leading computer scientist Salton, that meant giving up linguistics altogether! As indicated in this presentation, a view of language in accordance with Bakhtin (1981), and Kuhn (1962) is my working strategy.

For information science with KO philosophy of science is also directly important, as we may see in the concept "information", "relevance", and many other fundamental concepts – see further Hjørland (2021) & Hjørland (2024b).

I'll finish by indication that the important overall framework, "social epistemology" was first proposed by library and information scientist Jesse Shera (1951) in an article about classification. This is

something we should be proud of and continue to develop. It is about how knowledge is not primarily individual, but it is shaped in social and historical contexts – see further Hjørland (2024a).

5 Conclusion

Zeng *et al*. (2020) wrote:

> In categorizing and contextualizing vast amounts of developing information and by controlling or eliminating the pitfalls of semantic conflicts, KOSs ultimately underpin how experts and ordinary citizens understand the complexities of rapidly unfolding, global situations like those occurred during the COVID-19 pandemic (Zeng *et al*., 2020, p. 165)

Yes! In times of crises, the world depends on both:

a) new information structures specialized to the new challenges (such as new ontologies dealing with COVID-19 virus)
b) existing information structures being of high quality and updated with new information dealing with the new challenges (such as the International Classification of Diseases, ICD).
c) most important for KO as a field of research and practice is the focus on the common theoretical foundation of all kinds of KOS because this is what can keep us together.

Acknowledgments

Thank you to Lu An and Wuhan University for inviting me to give this keynote presentation and for the generous treatment I received.

References

ADAMS, Clive E.; POWER, K. Frederick A.; LEFEBVRE, C. An investigation of the adequacy of MEDLINE searches for randomized controlled trials (RCTs) of the effects of mental health care. **Psychological Medicine**, v. 24, n. 3, p. 741-748, 1994. Available at: https://doi.org/10.1017/S0033291700027896. Retrieved: 18 aug. 2025.

BAKHTIN, Mikhail. **The dialogic imagination**: four essays by M. M. Bakhtin. Austin: University of Texas Press, 1981. The book contains the chapter "Discourse in the Novel," originally written in Russian in 1934.

BENJAKOB, Omer; AVIRAM, Rona; SOBEL, Jonathan Aryeh. Citation needed? Wikipedia bibliometrics during the first wave of the COVID-19 pandemic. **GigaScience**, v. 11, p. 1-13, 2022. Available at: https://doi.org/10.1093/gigascience/giab095. Retrieved: 18 aug. 2025.

BENJAKOB, Omer; HARRISON, Stephen. From anarchy to wikiality, glaring bias to good cop: press coverage of Wikipedia's first two decades. *In*: REAGLE, Joseph; KOERNER, Jackie (ed.). **Wikipedia@20**: stories of an incomplete revolution. Cambridge, MA: MIT Press, 2019. p. 21-42. Available at: https://doi.org/10.7551/mitpress/12366.003.0005. Retrieved: 18 aug. 2025.

BRAMER, Wichor M.; GIUSTINI, Dean; KRAMER, Bianca M. R. Comparing the coverage, recall, and precision of searches for 120 systematic reviews in Embase, MEDLINE, and Google Scholar: a prospective study. **Systematic Reviews**, v. 5, n. 39, 2016. Available at: https://doi.org/10.1093/gigascience/giab095. Retrieved: 18 aug. 2025.

BRANNON, Brittany. *et. al*. Banned, challenged, questioned, or lost? Information access and the path to an information-resilient society: ASIS&T Midwest Regional Chapter. **Proceedings of the Association for Information Science and Technology**, v. 59, n. 1, p. 548-550, 2022. Available at: https://doi.org/10.1002/pra2.623. Retrieved: 18 aug. 2025.

BRIGANTI, Giovanni. How ChatGPT works: a mini review. **European Archives of Oto-Rhino-Laryngology**, v. 281, p. 1565-1569, 2024. Available at: https://doi.org/10.1007/s00405-023-08337-7. Retrieved: 18 aug. 2025.

CENTERS FOR DISEASE CONTROL AND PREVENTION (CDC). **COVID-19.** CDC, 2025. Available at: https://www.cdc.gov/coronavirus/2019-ncov/index.html.

CLAGGETT, Jennifer. *et al*. The effects of website traits and medical skepticism on patients' willingness to follow web-based medical advice: web-based experiment. **Journal of Medical Internet Research**, v. 24, n. 2, e29275, 2022. Available at: https://doi.org/10.2196/29275. Retrieved: 19 aug. 2025.

DE FREMERY, Wayne; BUCKLAND, Michael K.; LE DEUFF, Olivier. Resilient information science. **Proceedings of the Association for Information Science and Technology**, v. 59, n. 1, p. 559-560, 2022. Available at: https://doi.org/10.1002/pra2.626. Retrieved: 18 aug. 2025.

DELTELL, Luis; CLAES, Florencia. Free knowledge in times of pandemic. Study of articles on 'Covid-19' and the 'Covid-19 pandemic' on Wikipedia. (In Spanish). **Interface (Botucatu)**, v. 25, supl. 1, e200329, 2021. Available at: https://doi.org/10.1590/interface.200329. Retrieved: 18 aug. 2025.

DICKEY, Timothy J. *et al*. Accessibility, usability, inclusivity: information standards, guidelines, & best practices. Sponsored by the ASIS&T Standards Committee. **Proceedings of the Association for Information Science and Technology**, v. 59, n. 1, p. 561-564, 2022. Available at: https://doi.org/10.1002/pra2.627. Retrieved: 18 aug. 2025.

DRAGUSIN, Radu. *et al*. FindZebra: a search engine for rare diseases. **International Journal of Medical Informatics**, v. 82, p. 528-538, 2013a. Available at: https://doi.org/10.1016/j.ijmedinf.2013.01.005. Retrieved: 18 aug. 2025.

DRAGUSIN, Radu. *et al*. Specialised tools are needed when searching the web for rare disease diagnoses. **Rare Diseases**, 2013b. Available at: https://doi.org/10.4161/rdis.25001. Retrieved: 18 aug. 2025.

EUROPEAN CENTRE FOR DISEASE PREVENTION AND CONTROL (ECDC). **About ECDC**: what we do. ECDC, 2025.

GALLAGHER, John; LAWRENCE, Heidi Y. Rhetorical appeals and tactics in *New York Times* comments about vaccines: qualitative analysis. **Journal of Medical Internet Research**, v. 22, n. 12, e19504, 2020. Available at: https://doi.org/10.2196/19504. Retrieved: 18 aug. 2025.

GOOGLE. **Global digital compact.** Paper for the UN Secretary-General Guterres and participants of the Summit of the Future, dated 31 Mar. 2023. 2023.

HJØRLAND, Birger. *et al*. Some core concepts of domain analysis. **Em Questão**, v. 30, e-141781, 2024b. Available at: https://doi.org/10.1590/1808-5245.30.1417. Retrieved: 18 aug. 2025.

HJØRLAND, Birger. Information retrieval and knowledge organization: a perspective from the philosophy of science. **Information**, v. 12, p. 135, 2021. Available at: https://doi.org/10.3390/info12030135. An HTML version with adjusted bibliographical format was also published in ISKO Encyclopedia of Knowledge Organization.

HJØRLAND, Birger. Social epistemology. **Knowledge organization**, v. 51, n. 3, p**.** 187-202, 2024a. Available at: https://doi.org/10.5771/0943-7444-2024-3-187.

HONG, Yi; ZENG, Marcia Lei. International classification of diseases (ICD). **Knowledge Organization**, v. 49, n. 7, p. 496-528, 2022. Also available in: "ISKO Encyclopedia of Knowledge Organization" by Birger Hjørland and Claudio Gnoli (ed).

HORSNELL, Verina. The intermediate lexicon: an aid to international co-operation. **Aslib Proceedings**, v. 27, n. 2, p. 57-66, 1975. Available at: https://doi.org/10.1108/eb050493. Retrieved: 18 aug. 2025.

KAWAMURA, Keiichi. **Broad system of ordering.** Tokyo: Jusonbo, 2023.

KUHN, Thomas S. **The structure of scientific revolutions.** Chicago: University of Chicago Press, 1962.

LARDERA, Marco; HJØRLAND, Birger. Keyword. **Knowledge Organization**, v. 48, n. 6, p. 430-456, 2021. Available at: https://doi.org/10.5771/0943-7444-2021-6-430. Retrieved: 19 aug. 2025. Also available in "ISKO Encyclopedia of Knowledge Organization", by Birger Hjørland and Claudio Gnoli (ed).

LEFEBVRE, Carol. *et al*. Chapter 4: searching for and selecting studies. *In*: HIGGINS, Julian. *et al*. **Cochrane handbook for systematic reviews of interventions.** Version 6.4 (updated October 2023). London: Cochrane Collaboration, 2023.

LIÉVIN, Valentin. *et al*. FindZebra online search delving into rare disease case reports using natural language processing. **PLOS Digital Health**, v. 2, n. 6, e0000269, 2023. Available at: https://doi.org/10.1371/journal.pdig.0000269. Retrieved: 18 aug. 2025.

MANIEZ, Jacques. Database merging and the compatibility of indexing languages. **Knowledge Organization**, v. 24, n. 4, p. 213-224, 1997.

MAZZOCCHI, Fulvio. Drawing lessons from the COVID-19 pandemic: science and epistemic humility should go together. **Notes & Comments**, v. 43, article 92, p. 1-5, 2021. Available at: https://doi.org/10.1007/s40656-021-00449-9. Retrieved: 18 aug. 2025.

NEUHAUS, Fabian. Ontologies in the era of large language models – a perspective. **Applied Ontology**, v. 18, n. 4, p. 399-407, 2024.. Available at: https://doi.org/10.3233/AO-230072. Retrieved: 18 aug. 2025.

NÉVÉOL, Aurélie; DOGAN, Rezarta Islamaj; LU, Zhiyong. Author keywords in biomedical journal articles. **AMIA Annual Symposium Proceedings**, p. 537-541, 2010.

OBERHEIM, Eric; HOYNINGEN-HUENE, Paul. The incommensurability of scientific theories. *In*: ZALTA, Edward N. (ed.). **The Stanford Encyclopedia of Philosophy.** Fall 2018 Edition.

OPENAI. **ChatGPT**. 2022. Available at: https://chat.openai.com/. Retrieved: 18 aug. 2025.

OPENAI. **Introducing ChatGPT.** Blog. 2022. Available at: https://openai.com/blog/chatgpt. Retrieved: 6 jan. 2024.

PAPAJORGJI, Petraq J.; PARDALOS, Panos M. **Software engineering techniques applied to agricultural systems:** an object-oriented and UML

approach. 2nd ed. New York: Springer, 2014. Available at: https://doi.org/10.1007/978-1-4899-7463-1. Retrieved: 19 aug. 2025.

REYNOLDS, Heather L. *et al*. Green infrastructure for urban resilience: a trait-based framework. **Frontiers in Ecology and the Environment**, v. 20, n. 4, p. 231-239, 2022. Available at: https://doi.org/10.1002/fee.2446. Retrieved: 19 aug. 2025.

ROBERTSON, Stephen. Salton Award lecture: on theoretical arguments in information retrieval. **ACM SIGIR Forum**, v. 34, n. 1, p. 1-10, 2000. Talk at SIGIR 2000, Athens, July 2000. Available at: https://doi.org/10.1145/373593.373597. Retrieved: 19 aug. 2025.

SALTON, Gerald. A new horizon for information science. **Journal of the American Society for Information Science**, v. 47, n. 4, p. 333-335, 1996. Available at: https://doi.org/10.1002/(SICI)1097-4571(199604)47:4%3C333::AID-ASI13%3E3.0.CO;2-2. Retrieved: 19 aug. 2025.

SEGAULT, Antonin. Wikipedia as a trusted method of information assessment during the COVID-19 crisis. *In*: ROSSETTE-CRAKE, Fiona; BUCKWALTER, Elvis (ed.). **COVID-19, communication and culture**: beyond the global workplace. Oxfordshire: Routledge, 2022. p. 37-51. Available at: https://doi.org/10.4324/9781003276517-5. Retrieved: 19 aug. 2025.

SHERA, Jesse H. Classification as the basis of bibliographic organization. *In*: SHERA, Jesse H.; EGAN, Margaret E. (ed.). **Bibliographic organization**: papers presented before the Fifteenth Annual Conference of the Graduate Library School, July 24-29, 1950. Chicago: University of Chicago Press, 1951. p. 72-93.

SPANG-HANSSEN, Henning. How to teach about information as related to documentation. **Human IT**, n. 1, p. 125-143, 2001.

SPANG-HANSSEN, Henning. Kunnskapsorganisasjon, informasjonsgjenfinning, automatisering og språk [Knowledge organization, information retrieval, automation and language]. *In*: NORSK HOVEDKOMITÉ for klassifikasjon; Statens biblioteksskole; Norsk dokumentasjonsgruppe. Kunnskapsorganisasjon og informasjonsgjenfinning: seminar, 3.-7. desember 1973, Oslo. [**Workshop**]. Oslo: Riksbibliotektjenesten, 1974. p. 11-61.

SUNDIN, Olof; ANDERSSON, Cecilia; SÖDERSTRÖM, Kristofer. Controlling the machinery of knowledge: Google and access to COVID-19 grey-zone medicines. *In*: LUNDIN, Susanne; LIU, Rui; SMITH, Anja; MULLER, Elmi (ed.). **Medicine across borders**: exploration of grey zones. African Sun Media, 2023. p. 36-54.

WIKIPEDIA: The Free Encyclopedia. Category:COVID-19. *In:* WIKIPEDIA: the free encyclopedia. San Francisco, CA: Wikimedia Foundation, 2024a.

Available at: https://en.wikipedia.org/wiki/Category:COVID-19. Retrieved: 3 mar. 2024.

WIKIPEDIA: The Free Encyclopedia. Wikipedia and the COVID-19 Pandemic. *In:* WIKIPEDIA: the free encyclopedia. San Francisco, CA: Wikimedia Foundation, 2024b. Available at: https://en.wikipedia.org/wiki/Wikipedia_and_the_COVID-19_pandemic Retrieved: 3 mar. 2024.

WHITE, Howard D. Relevance in theory. *In*: MCDONALD, John D.; LEVINE-CLARK, Michael (ed.). **Encyclopedia of library and information sciences.** 4th ed. Boca Raton: CRC Press, 2018. v. 6, p. 3926-3939.

WORLD HEALTH ORGANIZATION (WHO). Infodemic. 2023. *In*: WORLD HEALTH ORGANIZATION. **Health topics**. 2023.

YOUSEFZADEH, Roozbeh; CAO, Xuenan. Large language models' understanding of math: source criticism and extrapolation. **arXiv**, arXiv:2311.07618, 2023. Available at: https://doi.org/10.48550/arXiv.2311.07618. Retrieved: 19 aug. 2025.

ZENG, Marcia Lei. *et al*. Implications of knowledge organization systems for health information exchange and communication during the COVID-19 pandemic. **Data and Information Management**, v. 4, n. 3, p. 148-170, 2020. Available at: https://doi.org/10.2478/dim-2020-0009. Retrieved: 19 aug. 2025.

ZENG, Marcia Lei; HONG, Yi. Challenges, opportunities, and approaches in a health KOS vocabulary's revisions – insights from the 11th revision of the International Classification of Diseases (ICD-11). **NKOS 2022 presentation proposal**, 2022.

Organização do conhecimento para a resiliência em tempos de crise: desafios e oportunidades

Resumo: Este artigo considera diferentes tipos de sistemas de informação, sistemas de organização do conhecimento e infraestruturas de informação em relação a problemas enfrentados em tempos de crise, como a pandemia de COVID-19. A análise dessas questões rapidamente conduz a debates mais fundamentais sobre a organização do conhecimento. Entre os sistemas discutidos estão infraestruturas gerais voltadas ao público em geral, como Wikipedia, Google e ChatGPT; estruturas específicas de domínio também destinadas ao público em geral, como sites relacionados à saúde; estruturas gerais para especialistas, como as bases de dados de citações Scopus, Dimensions, Google Scholar e Web of Science; e estruturas específicas de domínio para especialistas, como MEDLINE e FindZebra. Todos esses

sistemas são examinados sobretudo a partir da perspectiva da COVID-19. O artigo concentra-se em problemas comuns a todos os tipos de sistemas de organização do conhecimento, incluindo questões relativas a "recuperação de documentos" versus "recuperação de informação", "linguagem natural" versus "vocabulário controlado", "melhor correspondência" versus "correspondência exata", além de debates relacionados à filosofia da ciência. Conclui-se que, em tempos de crise, são necessários tanto novos sistemas especializados quanto a atualização dos sistemas já existentes. O aspecto mais importante para a organização do conhecimento, enquanto campo de pesquisa e prática, é o foco na base teórica comum a todos os tipos de sistemas de organização do conhecimento, pois é esse fundamento que pode manter o campo unido.

Palavras-chave: sistemas de organização do conhecimento; infraestruturas de informação; resiliência; questões teóricas

Authorship and Responsibility Statement

Conception and design of the study: Birger Hjørland
Data collection: Birger Hjørland
Data analysis and interpretation: Birger Hjørland
Writing: Birger Hjørland
Critical review of the manuscript: Birger Hjørland

Data availability statement

Not applicable.

Corresponding authorship

Birger Hjørland

birger.hjorland@hum.ku.dk

Editor-in-chief

Thiago Henrique Bragato Barros

How to cite

HJØRLAND, Birger. Knowledge organization for resilience in times of crisis: challenges and opportunities. **Em Questão**, Porto Alegre, v. 31, e-148670, 2025. https://doi.org/10.1590/1808-5245.31.148670

Receveid: 07/08/2025
Accepted: 08/04/2025

[1] Keynote presentation held at the Eighteenth International Society for Knowledge Organization Conference, Wuhan, China on March 20, 2024: Some sections of this paper were omitted by the oral presentation to accommodate to the time frame.

[2] A KOS is a system of concepts and selected semantic relations. Information infrastructures such as search engines and Wikipedia are not KOSs in themselves, but it is relevant to consider the theory of KOS also in relation to these infrastructures, as it is done in this presentation.

[3] The number of Google raters mentioned by Sundin *et al*., has not been verified and seems highly implausible. Other sources mention between 5.000 and 16.000 raters. The total number of employes and contractors in Google's parent company, Alphabet, was, according to a 2023 report, 121,000 contract workers and 102,000 direct employees.

[4] There is an article with the same title by another author and a different year, volume, issue and page numbers: DAHLBERG, Ingetraut. 2006. Knowledge Organization: A New Science? **Knowledge Organization**, v. 33, n. 1, p. 11-9, 2006. ChatGPT's reference is thus extremely erroneous although with some relations (title and journal name) to an existing reference.

[5] Since 2023-2024, when the article was written, ChatGPT version 5 has been launched, and among other things the following significant improvements have been made. (1) ChatGPT is no longer dependent on the data it was trained with, but can search for new data during a session (using "RAG"-like steps, each RAG stands for "Retrieval-Augmented Generation") and (2) ChatGPT can therefore verify its bibliographic references, and has improved these significantly. The main point of the article still applies, however: There are problems with the connection between the "information" systems bring/postulate and the basis of this information in the scientific literature.