Information Retrieval and Knowledge Organization: A Perspective from the Philosophy of Science

Birger Hjørland

Department of Communication, University of Copenhagen, 8 Karen Blixens Plads, DK-2300 Copenhagen S, Denmark; birger.hjorland@hum.ku.dk

Abstract: Information retrieval (IR) is about making systems for finding documents or information. Knowledge organization (KO) is the field concerned with indexing, classification, and representing documents for IR, browsing, and related processes, whether performed by humans or computers. The field of IR is today dominated by search engines like Google. An important difference between KO and IR as research fields is that KO attempts to reflect knowledge as depicted by contemporary scholarship, in contrast to IR, which is based on, for example, "match" techniques, popularity measures or personalization principles. The classification of documents in KO mostly aims at reflecting the classification of knowledge in the sciences. Books about birds, for example, mostly reflect (or aim at reflecting) how birds are classified in ornithology. KO therefore requires access to the adequate subject knowledge; however, this is often characterized by disagreements. At the deepest layer, such disagreements are based on philosophical issues best characterized as "paradigms". No IR technology and no system of knowledge organization can ever be neutral in relation to paradigmatic conflicts, and therefore such philosophical problems represent the basis for the study of IR and KO.

Keywords: information retrieval; knowledge organization; philosophy of science; classification; knowledge organization systems; ontologies; Kuhnian paradigm theory; pragmatism

1. Introduction

Information retrieval (IR) and knowledge organization (KO) are two research fields that, on the one hand, are separate fields of study, but on the other hand, have the same aim: to facilitate the findability of documents, knowledge, and information. Anderson and Pérez-Carballo's [1] handbook of KO has the title *Information Retrieval Design*, which indicates the close connection between KO and IR, where IR is about search processes, while KO is about designing optimal structures for IR (in addition to other purposes KO may serve).

Today, the field of IR has mainly migrated from information science to computer science and has developed systems used in search engines such as Google, which are extremely successful. There is, however, a surprisingly simple objection to the underlying principles in mainstream IR research: they are not based on scientific or scholarly norms on which documents or knowledge claims have the best scientific or scholarly foundation and correspond to our best theories and findings. It seems obvious that users need to retrieve what is regarded as true knowledge (or to be more precise, what is considered our best substantiated knowledge claims). The main approaches in IR are discussed in Section 3.

In this respect, the field of KO has often had the goal, at the least implicitly, to classify and represent documents and knowledge in accordance with updated scientific knowledge (e.g., to classify books about birds in accordance with the classification of birds in ornithology). Despite this aim, the theory of KO has lacked an adequate toolset of conceptualization to cope with it (and often, approaches have dominated that do not deal adequately with the problems. This is discussed in Section 2).

The purpose of this article is to provide arguments and conceptualization for creating methods and principles to develop systems and processes in IR and KO that aim at

providing users with our best substantiated knowledge. To do so, this article argues, we need to base IR on KO, and we need to base KO on insights derived from the philosophy of science.

2. The Field of Knowledge Organization (KO)

KO is sometimes termed "information organization" and [2] (p. 391) found that "information organization" is now by far the most used name for courses in educational programs in North American ALA-accredited programs. It is also sometimes termed "knowledge representation", as in the subtitle of the official journal of ISKO, Knowledge Organization: "International Journal devoted to Concept Theory, Classification, Indexing and Knowledge Representation". KO is about knowledge organizing processes such as indexing, tagging, classifying, describing, and organizing documents and information, and about knowledge organization systems (KOS) such as classification systems, thesauri, and ontologies. KO has a practical aim. Although it is also a theoretical field, the aim of both practical activities and empirical and theoretical research is to support practical activities.

The practical aims are the sole justification for KO. It has been claimed, for example, that KOS such as thesauri and controlled vocabularies are obsolete today, cf., [3]. KO must show that its processes and systems are necessary. If, for example, search engines such as Google can fulfill all practical needs without relying on KO, there is no need for this field. KO must always consider itself in relation to all possible alternative approaches, whether they have been developed inside or outside the field itself. Specifically, we must demonstrate, why, for example, Google is not enough (cf., [4]). This article contains arguments about the necessity of KO.

The KOS developed for organizing documents and information are primarily about organizing concepts. The conceptual structures used in KOS are to a large degree derived from specific domains of knowledge. Documents about birds, for example, are often organized according to how ornithologists organize the birds themselves. This principle was recognized by Henry Bliss [5], who argued (p. 37): "To make the [bibliographic] classification conform to the scientific and educational organization of knowledge is to make it the more practical", and this is probably the main reason the field was and still is named KO.

The roots of KO lie in the following:

(1) the practical classification and indexing of books and other kinds of documents in libraries and bibliographical databases;
(2) philosophical principles, including Aristotle's logic and Francis Bacon's classification of knowledge, among many others; (The reader should be warned, however, that there are many misunderstandings about philosophical issues, including Aristotle's role in classification. What is commonly attributed to Aristotle is a myth (see, e.g., [6], Chapter 2: "The Aristotelian Framework").
(3) scientific and scholarly contributions, including, for example, the contributions of Aristotle, Carl Linnaeus, Charles Darwin and many other scientists to the classification of living organisms and all other things in the world;
(4) developments in information technology (IT), such as databases, communication networks and social media.

The connections between these four fields are important for the development of KO as a field of study and practice but have often been neglected.

This article argues that subject knowledge and its foundation in the philosophy of science is the most important perspective for KO, but this perspective has so far not been very influential in KO. Alternatively, influential perspectives have been:

- To consider KO an intuitive process that does not need deeper justification. The original classification of journals in the citation indexes published by the Institute for Scientific Information, for example, was just intuitive (cf., [7], p. 602). In practice, many KO activities have been driven by this view: you simply start constructing a classification and stop when it seems to suit the purpose. A criticism of this perspective

may state that any classification is always serving some interests at the expense of other interests, and if unanalyzed, it cannot be optimized for the purpose it is going to serve (cf., [8]). Of course, this view challenges the whole idea of KO as a needed field of study.

- To claim that since only the individual user knows their own "information need", they are the only person qualified to set principles of what should be found in IR and how documents should be indexed or classified. This is the view underlying user-oriented and cognitive approaches in information science and KO and is discussed a little in the present article and more detailed in [9].
- To say that the best way, both economically and qualitatively, is to base KO on user tagging or related "social technologies", implying that a deeper theoretical understanding of KO is unnecessary. User tagging is by some both seen as "democratic" and economically preferrable, while there are also critical voices (for a discussion, see [10]).
- To claim that KO is basically a technological problem, that, for example, when enough computer power and optimal algorithms are available, the problems of IR will be solved without a deeper theoretical understanding of KO. A version of this view is that the principles underlying systems such as Google are sufficient. This view dominates the IR tradition in computer science and shall be discussed further in Section 2. (Computer science is, however, also dominating in knowledge representation and ontology development; therefore, a simple dichotomy between knowledge organization and computer science cannot be made.)

Approaches to KO have mostly been dominated by the dichotomy between psychological, cognitive understandings on the one hand, and technological understandings on the other. It has been claimed, for example, that there are two kinds of "relevance": human based relevance assessment and computer-based assessments. This dichotomy was criticized by [11] who argued that "systems relevance" is an oxymoron because systems are made by humans who have determined the criteria of relevance used by the systems. Relevance in the user-oriented school considers "real users" (as opposed to subject expert) to be best to evaluate the relevance of documents and information. In the next section, we consider a thought experiment of searching for cities in Sweden and here it is claimed that the relevant information may be found in a map or gazetteer constructed by some experts, not what the user believes are Swedish cities. Both the "systems approach" and the "user approach" (as well as the dichotomy itself) are therefore problematic positions, and a third way is needed.

A third way is the domain analytical approach [12], according to which both human information needs and technological approaches are understood as influenced by the understanding and background knowledge of the actors (including the computer programmers and mediators), which is shaped by the social and disciplinary contexts, the traditions, and the paradigms in which the actors have been socialized.

3. Challenges from IR

Search engines such as Google represent an impressing technology and its importance as aid to find relevant documents and information can hardly be overrated. We should acknowledge that both IR in general and search engines like Google have, in many ways, turned deeply rooted beliefs within library and information science and KO upside down. Nonetheless, everything has its limitations, and it is the job of research to suggest new ways forward. In this connection, it is important to consider the purposes of searching. Most users are probably not interested in exhaustive searches, but in high precision, and this may be one of the reasons for the popularity of Google. However, for some purposes, in particular academic purposes, exhaustive searches are often necessary, and it is important that tools are available for such purposes, and this may well be one of the weaknesses of Google and related systems.

When you search a system like Google, you typically type in a few words and study the first part of the result list (we shall not here consider, for example, picture or music

retrieval, but we consider the fundamental principles to be the same). This principle, that the system as a response to an input retrieves a set of documents, was by [13] termed "query transformation" and was by him opposed to a principle that is far older but less influential today, which he called "selection power," which is about the user's ability to make relevant distinctions during a search (these principles are further discussed in [14]). The principle of query transformation implies that you typically must know the words (or other symbols) that correspond to the words (symbols) in documents you would like to retrieve. This poses a theoretical problem, because it seems impossible to select terms from documents you do not know (since, following Socrates [15] if you already knew them, you would not be making a subject search for them). This has been interpreted as a claim about the principal impossibility for searching in general. In library and information science the distinction between "known item search" and "subject search" is well established. Sokrates/Plato's argument is clearly not about know item searching, which obviously is unproblematic. In relation to subject searching this seems also unproblematic, as everyone is doing this on Google every day. There is an important point, however. Unknown documents of relevance for a given inquiry may have to be searched in a context and with concepts as well as symbolic systems unknown for the searcher. This is in particularly the case if there have been paradigm shifts in the field of enquiry. Don Swanson is perhaps the only information scientists who have ever expressed the depth of this problem. He concluded [16] (p. 114): "Any search function is necessarily no more than a conjecture and must remain so forever".) The problem of knowing relevant search terms is, of course, made smaller because an initial search may provide hits containing further potential words to search (related to technologies known as "query expansion", which often depends partly on KOS for identifying synonyms, narrower terms, etc.). This means that iterative searches partly remedy the problem of identifying relevant search terms. Still, however, the initial conceptualization of the topic of the search is important. The main difference in relation to searches based on KOS is that the latter provides conceptual structures to help navigate and thereby identify relevant terms, symbols, and concepts.

A KOS may, for example, be a classification of cities according to geography, as done, for example, on a map or in a geographical classification system (a gazetteer), in which you are informed about a conceptual structure and may make adequate selections as you go. If you are interested in information about, for example, Swedish cities (including, perhaps, towns, villages and other settlements classified as such), this can be carried out about Sweden in general or about a region of Sweden; you do not have to know their names beforehand but can just use the classification.

This is probably the basic difference between mainstream IR and KO. In IR, you typically depend on a match between a search term and documents containing this term (in title, abstract, full text, etc. IR may of course also apply information such as descriptors, classification codes, etc., from KOS, but the main approaches in IR is about information from the document itself, not about value-added information. The utilization of reference lists in documents for IR mostly form part of the field of bibliometrics, rather than in mainstream IR (see [17,18]). The study of how different parts of both the documents themselves and information added in bibliographical records contribute as "subject access points" for IR is discussed by [19]). Contrasting to IR KO typically is about KOS containing semantic relations between concepts and providing, for example, a full list of cities in a particular part of Sweden. In IR, there is a technique known as "relevance feedback" in which users may indicate whether a found item is relevant or not, and the system may modify its search by including words from items that the user has marked relevant (a possible technique for "query expansion"). The search may also eliminate words from items marked nonrelevant, and thus increase the precision of the search. This technique still presupposes, however, that the searcher knows which words are relevant. In our geographical example, if a user does not know whether a given city name suggested by the system is about a Swedish or a Norwegian city, they may not be able to provide useful feedback, and the feedback may be harmful by making the system suggest Norwegian cities rather than Swedish. If a user

is searching for Swedish town, then "Stockholm" is a relevant hit, whereas "Oslo" is an incorrect hit. Therefore, criteria for what should be found (i.e., what is relevant) are not to be found in the searcher's belief (or in psychological studies or psychological studies as suggested by an influential school in information science). Criteria for what should be found must be based on KOS containing the best existing descriptions of the realities, often derived from scientific and scholarly studies. We recognize that the objectivity of science is an issue being debated in science studies. Still, however, this does not make any view as good as any other, and our geographical example demonstrates a relatively uncontroversial case.

Looking at the principle underlying Google's search engine, we find that four main principles are (a) "exact match", (b) "best match", (c) popularity measures, and (d) personalization (we shall not consider other issues such as the contents of the database and the influence of advertising, which are separate questions, less connected to the core theory of IR and KO).

(a) If you type in a sentence, such as: "It seems to be based on the problematic assumption that relations between concepts are a priori", Google will retrieve the one article (and its possible copies and versions) that contains this exact sentence. This exact match is obtained because the search applies proximity operators and thereby can retrieve documents identical to the query. This is not, however, what is generally understood by "exact match" techniques (or "set-retrieval"), which were defined by [20] (p. 284):

"Exact-match retrieval models use matching functions that, given a query, partition the document collection into two sets, those that match the query and those that do not. Documents in the matching set are generally not ranked, (although they may be ordered by date, alphabetically, or some other criterion. Exact-match models are generally simple and efficient and form the basis of most commercial retrieval packages [in 1992, not in 2021]. By far the most common exact-match model is the Boolean model". (It is no longer true that retrieved sets are not ranked. In 2019, "best match" replaced "most recent" as the default sort order for search results in PubMed, cf., [21,22]).

Exact match techniques allow searchers to use a "building blocks search strategy", in which well-defined search sets are constructed and combined by the searcher (see [23], p. 242). Exact match systems may or may not be combined with one or more KOS, but the principles behind exact match should be considered separately from the principles of KOS.

A disadvantage by exact match search techniques—if not combined with one or more KOS, as they usually are in "classical databases"—is, as we have seen, that the search is vulnerable about the users knowing the right terms, why the search may not be effective in terms of recall and precision. "Classical databases", such as MEDLINE, are based on exact match techniques, and this seems important for serious searches, where high recall is important (e.g., in evidence-based medicine, cf [14]). (Ref. [24] found that "the main problem of Boolean searching is not its performance. For many users the main obstacle is being able to use Boolean logic effectively in order to formulate queries in the way a commercial retrieval system requires".)

(b) If you type in a number of terms such as "concepts", "relations" and "a priori" from the example in point (a), Google will make a so-called "best match" search (also called "partial match", "relevance ranking" or "weighted retrieval") and retrieve millions of documents in a ranked order according to some principles, which are more or less business secrets. (The one article retrieved in point (a) is not among the top results).

To a large degree, the principles used are well established principles from IR research (see, e.g., [25–27]). These principles are mainly based on the relative frequency of terms in the whole database or collection, in the single documents and in the queries, in addition to issues such as the lengths of documents and the proximity of query terms within a document. Well-known examples of best match technologies are "vector space" [28] and "probabilistic" models [29]. Additionally, kinds of artificial intelligence (machine learning techniques) are used, in which algorithms learn to distinguish relevant documents and rank

them accordingly (see, e.g., [21]). Such technologies are considered superior in computer science. As [30] wrote: "statistical approaches won, simply. They were overwhelmingly more successful [compared to other approaches such as thesauri]".

The disadvantages of best match technologies are (1) that the searcher does not have full control over the research process (but saves time if they choose to trust the algorithms), and (2) that the system ranks documents according to "relevance", as if relevance is an objective concept, which clearly is a problematic assumption because different scientific perspectives and paradigms have different criteria for relevance (see [11,31]). Best match techniques are often based on similarity measures, but "similar" is a relative concept: Anything is similar with any other thing, depending on criteria, and it is easy to demonstrate that similarity based on words can be problematic because that means that a given text will by this criterion not be considered similar to its translation; (3) the techniques work on words or symbol structures (as opposed to concepts), which are associated with different meanings in different contexts. However, finding documents about a given subject is different from finding documents containing given words, and even containing given concepts (see [32] Section 3). Furthermore, because the principles behind best match are statistical, there is an element of popularity measure in this. If a given term is associated with a given meaning in a given context, that meaning cannot be separately identified; instead, the dominating meaning influences what is found. In other words, the principle disregards the insight from Kuhn that terms change meaning after scientific revolutions, or, said differently, in a way, it considers signs to be independent of context, which is a problematic assumption.

Today, best match techniques represent the dominating IR-paradigm, as applied in search engines. It challenges the techniques applied in "classical databases" such as MEDLINE. (Ref. [33] wrote: "The results from this study, then, support Robertson and Thompson's [34] conclusion that there is little difference in the levels of efficiency between weighted and Boolean retrieval mechanisms, but directly contradict the statements made by Belkin and Croft (1987) [35] and Turtle and Croft (1992) [20] about the superior performance of partial match techniques over exact match techniques. These results do not in any way prove the superiority of exact match techniques over partial match techniques, but they do suggest that different queries demand different retrieval mechanisms. Further studies and analyses are needed to determine which elements of a query make it best suited for partial match or exact match retrieval."). In evidence-based medicine (EBM), the use of classical databases still represents the dominating approach but is increasingly being challenged by best match techniques. [14] present arguments for set-based techniques for purposes for which high recall is important (such as EBM).

The change in the dominating search paradigm from selection power to query transformation implied a dequalification of professional searchers (information specialists) and of competent users. It is characteristic for professional searchers that they master a wide range of strategies to increase recall (finding more relevant documents) as well as to increase precision (avoiding more non-relevant documents). [36] (pp. 4–5; italics in original) pointed out, however, that with the best match technologies used by search engines, not only has this mastering disappeared, but even the concepts have lost their meaning:

"To make a more direct statement: the concept of a precision-enhancing device has one meaning in the context of set-based retrieval, and another and quite different meaning in the context of ranked-output retrieval. The 'same' device (such as 'using phrases') might very well be a precision device in one context and not in the other. The term *precision device* itself was coined in the former context: whether it is a valid concept for the latter is not obvious.

To put it even more directly, in accordance with the status of old IR hand which I suppose the Salton award means I must have acquired: Precision devices aren't what they used to be!" The same is the case for recall devices (p. 5) "Just to pursue the same line of argument a little further, recall devices are also problematic, despite their logical (rather than statistical) status in the set-retrieval context. Again, in the TREC tradition, we tend to measure recall at some arbitrary large cutoff (say 1000 documents). This immediately

destroys any claim to logical status for a recall- enhancing device. Even if we do something which (logically speaking) can only increase the size of the retrieved set, such as expanding the query with a lot of synonyms, it might still reduce recall at 1000 documents".

It seems that we, in the new context, have not only lost the professional ability to make qualified searching, but even in our research have lost important concepts and thereby part of the ability to understand what is going on in terms of optimizing IR.

(c)　It is well known that Google also uses a kind of popularity measure; the more in-links a certain document has, the greater the weight it is given and the higher it is listed in the ranked order shown to the user. This is often, if not in most cases, working very well because people often want the same as the majority. However, in searching for rare diseases, for example, this has proven a bad principle because rare diseases are, by definition, not a majority issue (for empirical demonstration of the failure of this principle for IR about rare diseases, see [37,38].

(d)　The fourth major principle in search engines is personalization; Google can identify users' IP address and thereby their physical location as well as their search history on Google, and may adapt, not just the advertisements, but also the so-called "organic results" in the ranked list presented to the user. This provides an element of subjectivity and randomness into the search and harms the ability to make conscious search strategies. It is also a double-edged sword; sometimes, it works well, but other times, you may want to eliminate this element, you may want more objective searches, you may have changed interests, or you may be searching on behalf of others. This is why the focus on past search interests may be harmful rather than fruitful.

A major problem in Google's principles is that scientific and scholarly criteria is absent in the four principles presented above. In our former example, we considered a search for all Swedish cities in a certain region of Sweden. Given a quality map, or gazetteer, such a query can be answered based on cartography or geographical research. This is just a simple example of the main view presented in this article, that what should be found by IR, and what should be represented in KOS, is what is considered true knowledge according to our best research and scientific theories about the content or subject matter. Although other examples may be more difficult than our geographical one, this does not invalidate the principle, but raises the philosophical problem about how science and scholarship obtain knowledge, how robust that knowledge is, and whether it reflects an objective reality. The principle that IR should find documents according to their scientific trustworthiness (rather than, for example, according to the searcher's guess on which words relevant documents must contain or according to popularity measures) seems to be an obvious demand, and it is surprising that this line or research in IR and KO seems to be almost entirely absent. Of course, academic quality and trustworthiness may be correlated with, for example, popularity measures and journal impact factors, but these are only indirectly associated with quality, and such correlations must be investigated before they should be relied on.

For now, we leave the question on how science and scholarship discover or construe knowledge, but shall consider how the scientific and scholarly quality of search results may be considered in IR. There are different strategies for doing so, including:

(1)　building specialized search engines rather than general ones (e.g., [37,38]).

(2)　selecting high-quality sources (e.g., journals with high impact factors); (See [18], Section 6 about the quality of indexed documents. Some citation indexes such as the Web of Science cover more limited amounts of indexed sources (based on journal impact factors), compared to, for example, Google Scholar, and use this to argue for a higher quality in the retrieved documents. This is, however, an open hypothesis, which seems to have been challenged by [39], who found that important papers are more and more published in non-elite journals. For a criticism of the journal impact factor, see, for example, [40]).

(3)　selecting documents based on principles used in so-called evidence-based research (e.g., studies based on double blind clinical trials); In evidence-based medicine (or evidence-based practice in general, EBP) the trustworthiness of claims about the

effectiveness of a given treatment are classified according to the quality of the research methods employed. Explicit norms should be made for investigations that are most relevant, and a hierarchy of the value of different kinds of research methods as evidence should be made (where randomized controlled trials are considered to be a high level of evidence, while, for example, evidence from expert committee reports is considered to be a low level of evidence). There has been criticism of such views, and there is an example of two different systematic reviews based on this procedure that provide very different conclusions (cf., [41]). Regarding IR, the EBP model provides clear criteria for prioritizing information sources, although, as already said, they are not uncontroversial.

(4)   Selecting documents based on their influence measures, e.g., their number of citations, in general or within some specifications (e.g., papers highly cited within leading journals in the field).

These possibilities are here only treated very shortly, while the focus in this article is about:

(5)   basing IR on quality KOS (such as our thought experiment with Swedish cities). In addition to such KOS, it is necessary that each document is assigned to the most relevant classes in the KOS, which is not a trivial issue, but depends both on the specific qualifications by the indexer and by the indexing philosophy used by the system, e.g., the operationalization of the concept "subject" (cf., [32]). (Ref. [42] Section 5.2, put forward the hypothesis, that indexing done by MEDLINE, one of the most important bibliographical databases in the world, may be based on principles that are too mechanical.)

To conclude this section, the working hypothesis behind this paper is that the dominating computer science approaches mainly focus on statistical relations between terms in single documents, terms in collections of documents and terms in queries, in addition to issues such as the lengths of documents and the proximity of query terms within a document. For example, a document that has all of the query terms occurring many times would be displayed first, followed by other documents where the query terms appear less. In addition, kinds of IR are used to identify "relevant" documents, often based on similarity measures. Although these approaches have so far been extremely successful, there is a need for alternatives based on knowledge of scholarly documentation and communication, including scholarly conceptions, traditions, and "paradigms" and the basis in the philosophy of science. This implies a more overall, top-down approach to IR. Such an approach is, among other things, about the construction of KOS, which is an interdisciplinary field in which computer science is also active, but which seems to be conflicting with dominating approaches to search engines.

4. Knowledge Organization Systems (KOS) and the Semantic Staircase

There are many kinds of KOS (cf., [43]). Here, we just look at three kinds: classification systems, thesauri, and ontologies. It is the hypothesis that KOS can be classified according to the semantic staircase (Figure 1), and that this model is the most important characteristics distinguishing kinds of KOS. Our sole aim in the description of these three kinds is to make this hypothesis clear to the reader. Therefore, many other issues concerning these KOS are not dealt with in the following presentation.
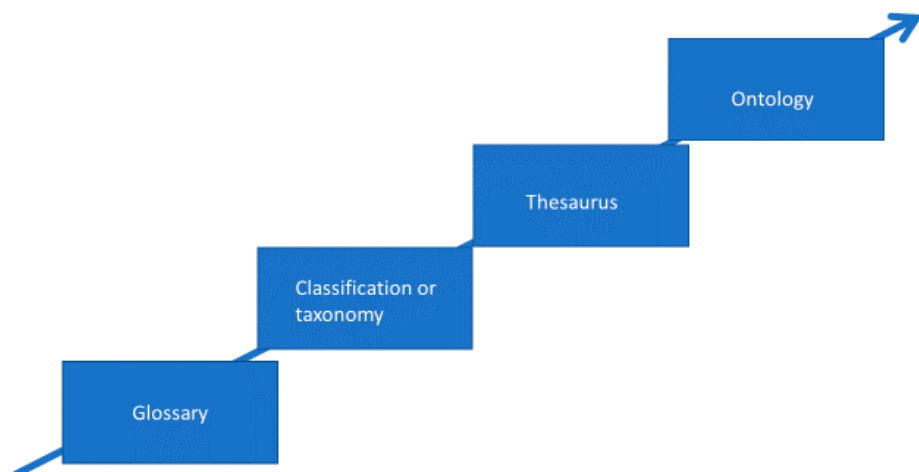
Figure 1. The Semantic staircase: Increasing semantic richness in kinds of knowledge organization systems (after [44]).

4.1. Classification Systems

Knowledge organization is primarily about bibliographic classification systems, which are systems for classifying documents and document representations. Documents are, however, about things in the world, e.g., about animals, celestial bodies, musical instruments, etc., and therefore bibliographical classifications to a high degree reflect classifications of things in the world. We may call such systems "scientific classifications" to distinguish them from bibliographic classifications, but "scientific classifications" in this sense include classifications made by scholars (e.g., musical instruments or literary genres) and "folk classifications" (e.g., genres of popular music). "Scientific classification" is here used as a generic concept for all kinds of classifications that bibliographic classifications may rely on. So, a bibliographic classification such as the Dewey Decimal Classification relies on how, for example, plants and animals have been classified by zoologists (although library classification systems in particular often are very conservative in updating, and unfortunately often reflect outdated knowledge, cf., [45]) (pp. 469–470). The present article is written on the assumption that to serve important functions for IR, KOS must be concerned with updated and trustworthy knowledge, rather than with outdated or problematic knowledge.

Figure 2 shows an example of a classification system. The intention of this classification example is to illustrate two points: The first point is that the dominating kind of relation in classifications is the generic relation (or "is a" relation); a bird is a vertebrate, and a vertebrate is an animal (although other hierarchical relations can also be used). The second point is that class names are also concepts: "birds", "vertebrates" and "animals" are classes as well as concepts. We may therefore say that a classification system organizes concepts (this claim is not generally accepted; however, the main spokesman for the view that KOS do not represent concept is Barry Smith, and we shall return to this point in Section 5). Finally, "animal", "vertebrate" and "bird" are also words (or terms or symbols), and concepts having verbal expressions are said to be lexicalized. KOS often, for example, list synonyms, which are relations between words rather than between concepts. Thus, classifications or KOS also display lexical relations between words (or, more broadly, between signs since other kinds of notations may be used). This is important because a major function of many KOS is to serve as controlled vocabularies.
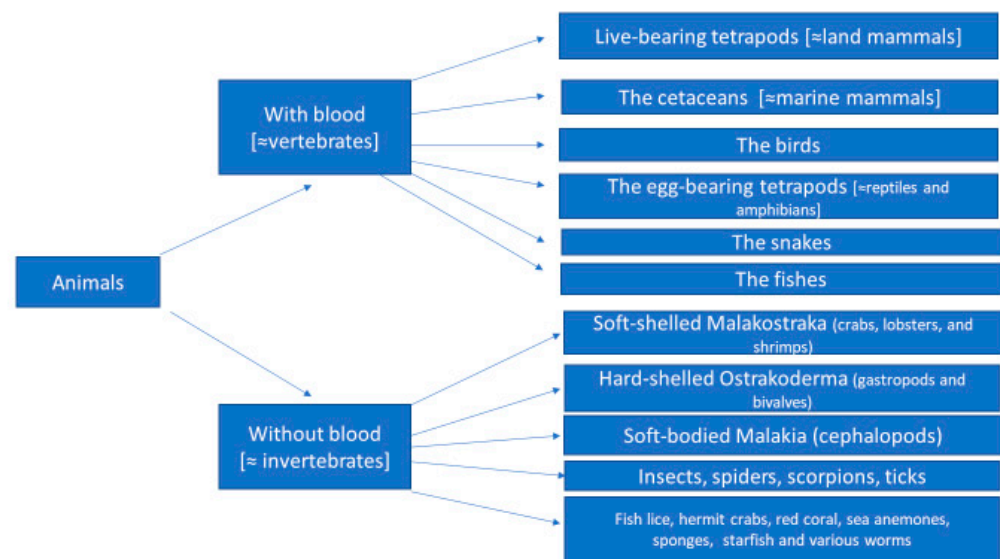
Figure 2. An example of a classification system.

4.2. Thesauri

Thesauri are defined in [46] (ISO 25964-1; clause 2.62) as: "controlled and structured vocabulary in which concepts are represented by terms, organized so that relationships between concepts are made explicit, and preferred terms are accompanied by lead-in entries for synonyms or quasi-synonyms".

A problem with this and related definitions in other standards is, however, that they fail to distinguish thesauri from ontologies. The solution suggested here is to define a thesaurus as a KOS with the limited and a predefined set of semantic relations it displays. If that set is expanded, we are no longer dealing with thesauri but with ontologies. Figure 3 shows an example from a thesaurus.

PEFERRED TERM R	Animals
BROADER CONCEPT	• Zoology
NARROWER CONCEPTS	• Aquatic animals • Birds • Domestic animals • Fauna • Insects • Laboratory animals • Mammals • Reptiles • Wild animals
RELATED CONCEPTS	• Animal behaviour • Animal diseases • Animal genetics • Animal nutrition • Animal resources • Cattle • Species
BELONGS TO GROUP	Science > Natural sciences
IN OTHER LANGUAGES	Animal French ЖивотныеRussian Animal Spanish

Figure 3. An example from UNESO Thesaurus [47].

A thesaurus may also contain a classification system, and this is the case with the UNESCO Thesaurus (the terms are then organized hierarchically rather than alphabetically). This also points to the narrow relations between kinds of KOS.

Explicitly, the ISO definition of thesauri [46] as well as the UNESO thesaurus organizes concepts. Mostly, thesauri also have lead-in terms (or synonyms) and thus also organize terms. We see again the generic relation between concepts (broader, narrower), but now also another relation: "related concepts". Thus, thesauri contain more explicit semantic relations than classification systems. (The UNESCO thesaurus is, by the way, not without problems. For example, zoology is normally not considered a broader concept in relation to birds, but zoology is considered a broader concept in relation to ornithology, and "vertebrates" and "animals" are broader concepts in relation to birds.)

4.3. Ontologies

Ontologies are kinds of KOS that are used in relation to front-end information technologies such as "the semantic web", but which are also used—in line with traditional KOS—for IR/literature searching (see e.g., [48]). Compared to classification systems and thesauri, they mostly have a much higher level of granularity and are closely related to actual scientific research in the domain. In practice, they tend to be more explicit and precise in their definitions. The Foundational Model of Anatomy ontology [49] describes the difference between an anatomical ontology and other anatomical tools such as atlases, textbooks, dictionaries, thesauri, or term lists. In relation to thesauri, it is written that: "Thesauri organize their content according to the meaning of their terms. However, since these terms are not explicitly defined, the meanings have to be implied by each user on the basis of perceived similarities and differences between terms. The FMA, by contrast explicitly defines the classes of its taxonomy, and links all these classes through an inheritance hierarchy to a single root: Anatomical Entity". Compare, however, this quote with the following by [50] (p. 34): "The symbol SN is used [in thesauri] to represent scope notes, which sometimes include definitions [. . .] As the trend is now toward the blurring of the difference between thesauri and terminological databanks, the tendency appears to be towards the increase in the number of definitions in thesauri. Svenonius [51] advocates the inclusion of as much definitional material as possible. More attention appears to be given to the form of definitions. For example, a model is suggested (Hudson, 1996) [52] for preparing logical definitions for indexing and retrieval thesauri".

The inclusion of more or less definitions should not be taken as a definitional difference, since this is a quality issue that varies widely among individual KOS. Ontologies are also mostly constructed by using formal languages, and many authors consider the use of such formal languages a necessary condition for a KOS to be an ontology.

Gruber [53] (p. 199) wrote: "A body of formally represented knowledge is based on a conceptualization: the objects, concepts, and other entities that are presumed to exist in some area of interest and the relationships that hold among them (Genesereth and Nilsson, 1987 [54]). A conceptualization is an abstract, simplified view of the world that we wish to represent for some purpose. Every knowledge base, knowledge-based system, or knowledge-level agent is committed to some conceptualization, explicitly or implicitly". Thereafter followed a widely cited definition:

"An ontology is an explicit specification of a shared conceptualization".

The same definition seems, however, to be valid for, for example, classification systems such as the periodic table in physics and chemistry. (Ref. [55] (p. 42) actually uses the periodic table as an example of an ontology and writes: "This example is interesting because it is an ontology with several facets of subclasses, and because it includes a system of instances as well as classes"). Gruber's definition also seems to fit biological taxonomies, and is therefore unspecific. It is a fundamental ambiguity in the use of the term "ontology" in information science in that it is both used (as in the semantic staircase Figure 1) as one specific kind of KOS with some specific requirements (see, e.g., [56]), and as a generic term for other kinds of KOS. Following the idea of the semantic staircase, it is here suggested to

define ontologies as kinds of KOS with the largest number of semantic relations between the concepts. It is typical of ontologies, however, that they often represent connections between entities and properties. The fish ontology architecture (Ref. [57] p. 6), for example, provides links between the property "extinct", the property category "fish status", and the entity "fish".

Another aspect of Gruber's definition is the already mentioned controversy about whether "concepts" and "conceptualizations" should be considered the units in ontologies, a problem that will be discussed later in this article. One issue related to this problem should, however, be mentioned here. Ontologies are often supposed to be tools from which any new tools needed in the future might be produced. Ref. [58] (p. 59) wrote: "The Foundational Model of Anatomy (FMA) ontology is being developed to fill the need for a generalizable anatomy ontology, which can be used and adapted by any computer-based application that requires anatomical information". They also wrote, however, that FMA "is both a theory of anatomy and an ontology artifact". As such, it is a conceptualization in Gruber's sense, and as such its application has inbuilt limitations in relation to applications based on other conceptualizations, an issue we come back to in the discussion of concepts and realism.

Ref. [59] suggested the following relations as the most important ones for biomedical ontologies:

- is_a
- part_of
- located_in
- contained_in
- adjacent_to
- transformation_of
- derives_ from
- preceded_by
- has_participant
- has_agent

We observe the waste expansion of semantic relations compared to classification systems and thesauri. There seems to be no limit to the number of relations that may be used, and new kinds are discovered (or constructed) when ontologies are constructed for new domains. Ontologies therefore differ from classifications and thesauri by providing many more kinds of semantic relations between concepts.

(An anonymous reviewer suggested that this list by Smith et al. [59] "is just making up reality, an illusion of reality for the sake of efficiency and consistency". Perhaps it is better to say that it is one selection among other possible selections, the fruitfulness of which represents a theoretical assumption in need of justification.)

4.4. The Semantic Staircase

The semantic staircase is a classification of KOS according to how many kinds of semantic relations they display. According to Olensky (2010), the term "semantic staircase" was first used in German by [60] Blumauer and Pellegrini (2006, p. 16; as "Semantische Treppe"). [61] (pp. 30–35) presented a related term "The ontology spectrum" which is a classification, not of KOS, but of "ontologies", which considers, for example, classification systems and thesauri as kinds of ontologies.

In our descriptions from classification systems over thesauri to ontologies, we showed that the number of semantic relations did indeed increase. Regarding Figure 1, it should be noted that a glossary typically is an alphabetic listing that does not provide semantic relations between the concepts, and that we consider "taxonomy" and "classification" to be synonyms.

The verification of the hypothesis of the semantic staircase requires that it can be shown that the higher forms of KOS are able to represent the lower forms. More research on this is needed, but some supporting views are:

Ref. [62] suggested that topic maps (which he says are based on an ontology framework) can represent other kinds of KOS:

"The relationship between topic maps and traditional classification schemes might be that topic maps are not so much an extension of the traditional schemes as on a higher level. That is, thesauri extend taxonomies, by adding more built-in relationships and properties. Topic maps do not add to a fixed vocabulary but provide a more flexible model with an open vocabulary. A consequence of this is that topic maps can actually represent taxonomies, thesauri, faceted classification, synonym rings, and authority files, simply by using the fixed vocabularies of these classifications as a topic map vocabulary".

Another supporting view is the idea that it is possible to transform a classification system into a thesaurus (without related terms). Ref. [63] examined if the second edition of the Bliss Bibliographic Classification can be used as a source for thesaurus terms and structure (i.e., concepts and their semantic relations). Although she described some problems, these problems seem not to refuse the idea that a classification can be transformed into a thesaurus, but just to point out some issues in the specific classification used (one could say some problems in its quality).

A third supporting view is that it seems to be evident, that if it can be established, for example, that a bird is a vertebrate (or in general that X is an A), then this is a building block that can be used in both classifications, in thesauri and in ontologies, or in any other kind of KOS. By implication, it seems unfruitful to continue studying, for example, classifications and thesauri, as separate objects. We should rather be studying KOS in general, i.e., concepts, conceptual systems, and conceptual/-semantic relations. (Nevertheless, of course, anything we have learned about, for example, classification schemes and thesauri, is still useful in this generalized context).

Different kinds of KOS are different in relation to the kinds of semantic relations they can display. However, a central issue is not about kinds of KOS but about determining the relevant/fruitful/true relations. An anonymous reviewer wrote: "In fact another interpretation [. . .] could be that ontologies impose the constraints (sometimes simplifications) required by computer logic and the definition of classes, inheritance and hierarchy that are proper of object-oriented programming, while the scope notes and the "logical inconsistencies" of thesauri could present a more flexible opportunity to explain and represent the complexities of science in a more humanistic/verbal way". This quote raises some deep and very interesting problems. In this connection it can be mentioned that [64] wrote about "Questioning the Univocity Ideal" in terminology studies and that [65] from a general philosophical point of view, wrote about praising vagueness. Such references support the question addressed by the reviewer, which, to my knowledge, never has been seriously addressed in knowledge organization. If fruitful, this view challenges the idea of the semantic staircase.

There are also more traditional issues about scientific and scholarly criteria for IR and KO. For example, how do we determine whether X is a kind of A? It has been claimed by [66] (p. 583) that "paradigmatic relationships are those that are context-free, definitional, and true in all possible worlds". At the same time, the literature demonstrates a common understanding that paradigmatic relations are the kinds of semantic relations used in thesauri and other knowledge organization systems (including equivalence relations, hierarchical relations, and associative relations). It is a strange claim. Relations between species of birds, for example, are, as discussed in Section 5.2 determined in ornithology and subject to different theories and paradigm shifts (see further [67].) To a large degree, this must be done, of course, by scientific research. So, a good KOS is one that is based in solid subject knowledge, while a bad KOS is not. The view expressed by [66] discourages information specialists to consider the scholarly literature. If KOS shall be used and serve their functions, they must be trusted by the users. Unfortunately, there are indications that this need not be the case (cf., [68] (p. 511) who wrote: "Over half the time searchers neglected to consult a thesaurus, they did so either because they did not trust the quality

of the thesaurus, because the thesaurus was not available, or because they had to search several databases for a request".)

5. Concept Theory and Realism

So far, we have assumed that concepts are the elements in KOS, but also said that this view is not without its detractors. The view that KOS are primarily about the organization of concepts and that concepts are units of knowledge, has been expressed by, among others, [69–72]. There are, however, reasons here to reconsider these claims. As formerly indicated, Barry Smith is probably the leading detractor, but also [73] asked if KO can do without concepts and Herre [74], although disagreeing with Smith by considering concepts important, has also saw universals as a necessary category in KOS. As we consider this discussion of high importance, we shall address it in some depth. The questions to be addressed in the subsections of this section are: Section 5.1: Challenges from "Smithian realism", should concepts be replaced by "universals" as the units in KOS? Section 5.2: What are concepts? The semiotic triangle. Are concepts units of knowledge? Section 5.3: Does a KOS need to contain universals and symbolic structures in addition to concepts as claimed by Herre [74] (p. 301)? Finally, in Section 5.4: Pragmatic realism.

5.1. Challenges from "Smithian Realism"

Smith and co-authors [75–77] discuss ontologies and criticize the conceptualist understanding. [77] writes (p. 7):

"The code assigned to France, for example, is ISO 3166-2:FR and the code is assigned to France itself—to the country that is otherwise referred to as Frankreich or Ranska. It is not assigned to the concept of France (whatever that might be)".

This example is somewhat atypical for concepts and a bit difficult because (1) France is an individual concept rather than a general concept; (2) the meaning of the concept "France" is determined by publicly available conventions: we know how the borders and the definition of a particular country is decided—typically following wars (whereas scientific concepts typically are developed by research). Nonetheless, even in this case, it can be argued that the ISO code 3166-2:FR is assigned to a part of reality determined by a conception, as we shall see when we introduce the semiotic triangle in Section 5.2. (The related term "Europe" was by [78] considered to represent many concepts.) We shall also consider concepts as dynamically changed to cope with the problems of their use. A concept such as France has been changed, e.g., by defining its territory in terms of maritime boundaries and airspace, when fishery, oil interests and airplanes made this important. Works and classification systems on "France" may vary in what is included and what is excluded under that term or symbol (they need not follow the legal concept. The Danish library classification system DK5, for example, in class 46, Denmark, includes some former Danish positions in the subclass 46.8 and thereby conflicts with the legal definition of Denmark). However, even the legal concepts "Denmark" and "France" are interpretations or conceptualizations that may be challenged in courts.

As the alternative for concepts, Smith, in many writings, has argued that "universals" or "types" are the units in KOS. [79] presented some definitions of realism, antirealism, nominalism and conceptualism, but then declared (pp. 140–141):

"Since 2002 we have been attempting to move beyond such disputes by developing a methodology, which we call 'ontological realism', that will capture what we believe to be a kernel of practical significance in these debates by addressing the question what it is to which the terms used in ontologies should be seen as referring. Because ontological realism is a methodology, and not a doctrine, it stands in no logical relation to any of the metaphysical doctrines specified above".

This capitulated attempt to base the realist methodology in deep philosophical arguments looks like a weakness; is it perhaps a partial retreat from earlier realist claims?

Types and universals are explained in this way (Ref. [79] (p. 141): "Types or universals— we shall always use these terms synonymously in what follows—are to be understood as

counterparts in reality of (some of) the general terms used in the formulation of scientific theories. Particulars are concrete individual entities (entities that exist in space and time and that exist only once); types or universals are to be understood as repeatable. This means that, for each given type, we can in principle discover of indefinitely many particulars that they are its instances").

It has been agued by Herre [74] that Smith and followers are wrong, that KOS cannot do without concepts. He wrote (p. 303) that Smith's position (described in, among other places, [75,76]) is:

- Universals have an observer-independent objective existence; they are invariants of reality.
- Bad ontologies are those whose general terms lack a relation to corresponding universals in reality, and thereby also to corresponding instances.
- Good ontologies are representations of reality. A good ontology must be based on universals instead of concepts.

Herre (pp. 303–304) found that the problem in Smith's argument is that in relation to condition 3, no definition for reality representation is provided and that there is no representation of reality without concepts. This is not seen as a problem for realism in general, only for what he [74] (p. 303) termed "Smithian realism" in contrast to his own "integrative realism".

Arp, Smith and Spear [77] (p. 7) further wrote:

"The goal of ontology for the realist is not to describe the concepts in people's heads. Rather, ontology is an instrument of science, and the ontologist, like the scientist, is interested in terms or labels or codes—all of which are seen as linguistic entities—only insofar as they represent entities in reality. The goal of ontology is to describe and adequately represent those structures of reality that correspond to the general terms used by scientists".

This quote is important because KOS should be based on knowledge developed by scientist and not—as assumed by cognitive theories—by psychological studies of people, which clearly seems a kind of idealism and opposed to realism. Still, however, we need to consider concepts. There are two levels involved in discussing concept in KOS: (1) the researchers doing the research are influenced by different paradigms, which make scientists' use of general terms concepts, rather than passive registrations of reality; (2) at the level of KOS-construction, the information scientists must consider the scientific literature. Again, this cannot be just as passive registrations of reality, selections and decisions are made. All this is missing in "Smithian realism", but is well-captured by [80] (pp. 32–33), who made a description for the real-life problems interpreting the scientific literature for metadata construction:

"TAIR curators wishing to collect data on a given gene (say the Unknown Flowering Object [UFO] gene in *Arabidopsis thaliana*) could not compile data from each relevant publication, as it would be too time consuming: even just a keyword search on PubMed for "UFO Arabidopsis" results in over fifty journal articles, only one or two of which are used as a reference for an annotation. Hence curators chose what they saw as the most up-to-date and accurate publications, which as a consequence became "representative" publications for that entity [. . .] These choices are impossible to regulate through fixed and objective standards. Indeed, bioinformaticians have been trying to automate the process of extraction for years, with little success. The very reasons why the process of extraction requires manual curation are the reasons why it is hard to divorce it from subjective judgment: the choices involved are informed by a curator's expertise and her ability to bridge between the original context of data production and that of data dissemination".

To conclude this Section: "Smithian realism" represents important arguments about finding the warrant of KOS in the scholarly literature rather than in the heads of people. Still, however, its insistence on "universals" and a naïve approach to reality representation is problematic; we cannot do without concepts.

5.2. What Are Concepts?

To understand concepts, we shall first have a look at the semiotic triangle (or "the triangle of meaning"). The version used here was produced by [81] (p. 14) in a discussion of the relations between thoughts, words, and things. The triangle has, however, a long history (see [82]) (pp. 58–59). The "referent" in Figure 4 is in some versions of the semiotic triangle called "object", "symbol" is sometimes termed "sign vehicle", and "thought or reference" may be termed "concept" or "sense". The broken line at the base of the triangle indicates that there is not a direct relationship between the symbol and the referent: A given symbol is only related to a referent by somebody (or a system) knowing this sense of the symbol (e.g., the term "cat" is only understood as referring to a cat by people or systems with this knowledge of English). The Danish structural linguist Louis Hjelmslev [83] (p. 81) demonstrated that a term in one language often does not correspond exactly to the equivalent term in another language. "Cat" may have a slightly different meaning in relation to the German "Katze": Each natural language is a classification system that classifies the world differently. Another way to express this is that the model involves mediation: the object is related to the sign vehicle via the mediation of the concept/sense. This understanding of the mediated nature of terms is in opposition to the view presented above as "Smithian realism" and to the corresponding suggestion to use universals as units in KOS. The mediated view relates to a broad range of philosophies. It is implied, for example, in Kant's philosophy, in Peirce's and other pragmatists' thinking, in hermeneutics (in the hermeneutic circle), and in Thomas Kuhn's philosophy of science, to which we now turn.
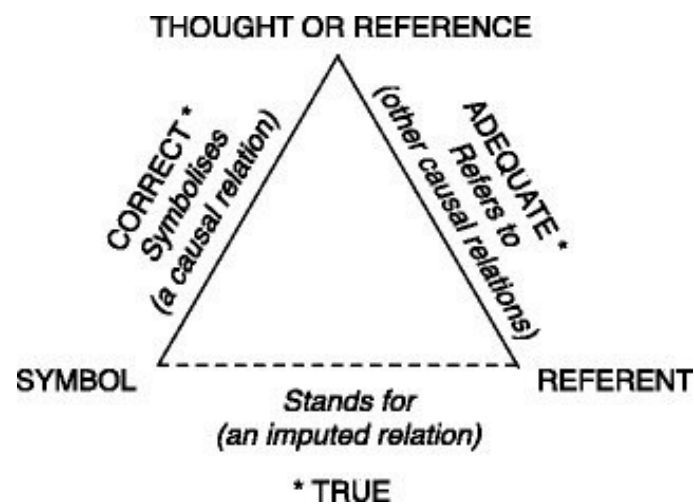
Figure 4. The Semiotic triangle [81] (p. 14).

Kuhn [84] wrote about scientific revolutions, for example, how the Copernican theory that the Earth revolves around the Sun replaced the former theory of Ptolemy that the Sun revolves around the Earth. Kuhn found that in such "paradigm shifts", the terms involved in the theories gained new meanings. Kuhn highlighted the issue of conceptual changes in science and has had a huge influence on the thinking about concepts and conceptual changes. [85] (pp. 666–667) wrote:

"[T]he acceptance of the Copernican theory that the earth revolves around the sun required the rejection of the Ptolemaic theory that the sun revolved around the earth. Replacement was not merely a matter of one theory being substituted for another, but also involved shifts in meaning of the concepts used in the theories. In the Copernican revolution, for example, the concept "planet" shifted to include the earth and exclude the sun and moon".

Kuhn thus exemplified changes in the meaning of terms:

Paradigm one: Ptolemaic astronomers might learn the concepts [star] and [planet] by having the Sun, Moon, and Mars pointed out as instances of the concept [planet] and some fixed stars as instances of the concept [star].

Paradigm two: Copernicans might learn the words "star", "planet", and "satellites" by having Mars and Jupiter pointed out as instances of the concept [planet], the Moon as an instance of the concept [satellite], and the Sun and some fixed stars as instances of the concept [star].

Thus, the terms "star" and "planet" gained a new meaning and astronomy gained a new classification of celestial bodies. A given KOS may organize concepts such as [star] and [planet] and such an organization necessarily represents a given paradigm. There is no way to define or organize "star" and "planet" independent of paradigm/interpretation. This example fits the semiotic triangle very well: there are no direct connection between a part of reality and the symbol used for it, the connection depends on the concepts, which again depends on conceptions/paradigms.

Some readers may argue that this example is not of much relevance because today we know that Copernicus was right. This is established knowledge and is no longer just considered a theory. Therefore, contemporary science need not consider [star] and [planet] as concepts based on a conception or a paradigm. We can just follow Smith and say that we are referring to universals or types understood as observer-independent invariants of reality. This is wrong, however (although typical for what Kuhn termed "normal science"). Contemporary science depends also on concepts that are based on underlying conceptualizations, and it is important to understand concepts as embedded in theories. Only one example shall be given here:

In 1992, the first volume of the Handbook of the Birds of the World [86] was published and volume 17, the most recent volume, was published in 2013. During the 21-year period of publishing this handbook, the classification of birds changed substantially (cf., [87]). There are indications that the shift in bird classification suggested by Fjeldså in the last volume represents a paradigm shift in the sense of Kuhn compared to the classification used in the work, decided before the first volume was published.

This shift of classification in ornithology is part of broader paradigm of developments in biological taxonomy during the 20th. Century. However, in opposition to Kuhn's description in the physical sciences, these biological paradigms tend to exist in parallel. This issue, whether paradigms only exist one at a time or at the same time as competing programs, is a point where many philosophers disagree with Kuhn. This is also the case with the present author, as reflected in definition of concepts as cited in the present article. Among these biological paradigms are "numerical taxonomy" (or phenetics), which use the overall likeness of organisms as a classification criterion and "genealogical classification" which use common ancestry as the basis of classification. The development of molecular techniques in biological taxonomy has of course also had a huge influence the development towards the new classification suggested by Fjeldså. The question for us is: has this revolution in classification changed the concepts used in ornithology, e.g., in the name of species and families of birds? [88] (p. 12; italics in original) provided the following example:

"When reviewing the manuscript of an excellent new ornithological text, one of us was startled to read the comment 'the Yellow-throated Longclaw (*Macronyx croceus*), a member of the pipit family (Passcridae).' Every ornithologist knows that the pipit family is the Motacillidac and that the Passeridae (the Old World sparrows), if recognized as a separate family, is composed of the sparrows of the genera *Passer*, *Petronia*, *Carpospiza* and *Montifringilla*. After recovering from our shock, we realized that the authors of this text were following the newly proposed classification by Sibley & Ahlquist (1990) [89], in which the pipits and their relatives are placed as a subfamily, Motacillinac, together with the Passerinac, Prunellinac, Ploccinac and Estrildinan, in an expanded family Passeridac. The point is not whether Sibley & Ahlquist are correct in allying these taxa into a single family but rather a matter of case of communication among all ornithologists. Even as specialists

in avian systematics, we had difficulty understanding the quoted expression and required several minutes to realize what the authors meant".

One of the characteristics of Kuhn's paradigms is the taxonomic incommensurability thesis that makes communication difficult between researchers working in different paradigms. Therefore, the quote by Mayr and Bock [88] can be taken as an indication that a Kuhnian paradigm shift has taken place in ornithology, and that the concepts indeed changed.

What, then, are concepts?

Whereas concepts are often understood as internal representations in individuals, Kuhn's theory considers their public, social nature as primary. Only secondarily concepts become internal representations as something individuals learn by participating in a culture, subculture, disciplinary tradition, or paradigm. On this basis, [90] (pp. 1522–1523) suggested the following definition:

"Concepts are dynamically constructed and collectively negotiated meanings that classify the world according to interests and theories. Concepts and their development cannot be understood in isolation from the interests and theories that motivated their construction, and, in general, we should expect competing conceptions and concepts to be at play in all domains at all times".

This definition partly provides an answer to the problem formulated in the start of this Section 4: are concepts units of knowledge? Kuhn's claim that concepts have different meanings in different paradigms can be considered a version of "meaning holism" or "semantic holism". The concept "information" seems to be a clear example in library and information studies (LIS), where, for example, Shannon's information theory, the cognitive view, semiotic and social-oriented views seem to provide incommensurable concepts of "information". We may still say that concepts are units of knowledge in the sense that any sentence expressing knowledge (or knowledge claims) are built on terms representing concepts. In relation to concept analysis, however, it seems necessary to consider the theories of which concepts form parts (e.g., one cannot answer the question "what is information" without considering the different theories and interests for which this concept is considered important).

5.3. Does a KOS Need to Contain Universals and Symbolic Structures in Addition to Concepts?

Herre [74] (p. 301) wrote: "We hold that any reasonable foundational ontology must include these three types of categories [universals, concepts, and symbol structures]". Because ontologies are kinds of KOS (cf., [91]), we shall here not limit the discussion to (foundational) ontologies but consider this problem as a general issue for all kinds of KOS. Herre's discussion is theoretically advanced, but nonetheless, there is a need to consider his definitions briefly. He defined these concepts as follows:

- Concept [74] (p. 301–302): "Concepts are categories that are expressed by linguistic expressions and which are represented as meanings in someone's mind. Concepts are a result of common intentionality which is based on communication and society (Searle, 1995) [92]". (Ref. [74] (p. 302, footnote 7): "The mental representation of a concept allows us to understand a linguistic expression. Concepts are outside of individual minds, but they are anchored, on the one hand, in individual minds by the concepts' mental representation, and on the other hand, in society as a result of communication and usage of language".)
- Category [74] (p. 301): "Categories are entities that are expressed by predicative terms of a formal or natural language that can be predicated of other entities. [. . .]. We distinguish at least three kinds of categories: universals, concepts, and symbol structures. We hold that any reasonable foundational ontology must include these three types of categories".
(It is difficult to understand Herre's difference between categories and concepts. Both categories and concepts may be expressed by linguistic expressions and predicative terms and may be represented in somebody's mind. Ref. [93] wrote: "Categories are

hard to describe, and even harder to define. This is in part a consequence of their complicated history, and in part because category theory must grapple with vexed questions concerning the relation between linguistic or conceptual categories on the one hand, and objective reality on the other". Concepts and categories are often defined in ways that make them synonyms, but categories may also, and probably better, be used about the highest kinds or genera, such as Aristotle's 10 categories: substance, quantity, quality, relation, place, date, posture, state, action, and passion. This way of understanding categories has its own philosophical history (see, e.g., [94]). For more detail about the relation to Ranganathan's categories in KO, see [95])

- Symbol/symbol structure [74] (p. 302): "Symbols are signs or texts that can be instantiated by tokens. There is a close relation between these three kinds of categories: a universal is captured by a concept which is individually grasped by a mental representation, and the concept and its representation is denoted by a symbol structure being an expression of a language. Texts and symbolic structures may be communicated by their instances that a[re] physical tokens". Further (p. 304): "One must distinguish between symbols and tokens. Only tokens, being physical instances of symbols, can be perceived and transmitted through space and time".
  Ref. [96] (p. 120): "Tokens are said to instantiate types: they exemplify, embody, manifest, fall under, belong to types; they're occurrences, instances, members of types. Tokens are treated as individuals, singles, particulars, substances, objects; they're concrete, real, material. Types, on the other hand, are like sorts, kinds, forms, properties, classes, sets, universals; they're said to be abstract, ideal, immaterial."

- Universal [74] (p. 301): "Universals are constituents of the real world, they are associated to invariants of the spatio-temporal real world, they are something abstract that is in the things".

A few objections to Herre's definitions are:

(1) It seems unnecessary to define categories and concepts as entities that are expressed by terms or linguistic expression. We might conversely say that concepts may be expressed by words, and that concepts that have a linguistic or symbolic expression are lexicalized. Ref. [97] (p. 237) wrote that WordNet introduced the non-lexicalized concept "wheeled vehicle": "The argument is that people distinguish between the category of wheeled vehicles and vehicles moving on runners independently of whether this distinction is lexically encoded in their language".

(2) As already presented in Section 4.1, there is a discussion in the ontological literature between a realistic position that rejects concepts as units for KOS and another realist position (like Herre and the present author), which defends concepts as units in KOS. Herre's statement that concepts are "represented as meanings in someone's mind" may, however, make it more difficult to see his position as representing realism (although his argument is partly saved by his addition that concepts are a result of a common intentionality, that they are social). We shall return to this problem about realism in Section 5.4.

(3) Herre finds that ontologies/KOS must include universals, but his position on this point seems not to be crystal clear. On the one hand, he writes [74] (p. 305): "In sum, the nodes in an ontology are labeled by terms that denote concepts. Some of these concepts, notably natural concepts, are related to invariants of material reality". This is in line with the semiotic triangle and in accordance with the view expressed by the present author. However, he [74] (pp. 326–327) also speaks of facts, defined as "The simplest combinations of relators and relata". Perhaps Herre's view is opposed to that of [98] (p. 4): "In hermeneutics, we defend the idea that there are no pure facts. Behind every interpretation lies another interpretation. We never reach an understanding of anything that is not an interpretation". If we follow this view, it seems that universals are not elements in ontologies since we can only know them as interpretations and concepts. This is further discussed in Section 5.4.

Words, terms, symbols, and symbolic structures are often used in KOS to represent concepts. In addition, KOS may contain other words that do not refer to concepts but just to other words. A KOS can be a pure conceptual structure (e.g., classifying animals as in Figure 2) and the relation between concepts are semantic relations. However, very often, a KOS is also constructed as a controlled vocabular, which includes synonyms for each concept. The concept [city] may be represented by, for example, the terms "city" and "metropolis". In thesauri, synonyms are displayed as "lead in terms" to the "preferred terms". We may distinguish between the pure conceptual or semantic structure itself (e.g., the kinds of animals and their relations) and the lexical (or lexical semantic relations) between terms, e.g., which terms are considered synonyms. The distinction made here between semantic relations and lexical relation is not common in the linguistic literature, which seems to be unclear on this point. This distinction represents, however, a tradition in knowledge organization. [99] (pp. 2–3), for example, wrote "WordNet makes the commonly accepted distinction between conceptual-semantic relations, which link concepts, and lexical relations, which link individual words". And: "Synonymy is, of course, a lexical relation, and terminology relations are relations between words or between words and concepts, but not relations between concepts".

In this context, it should be emphasized that synonymity is not an intrinsic quality of words but is the characteristics that two or more signs are considered replaceable/interchangeable by somebody, in a given context. An example: *The Thesaurus of Psychological Index Terms* (visited 8 February 2021) lists "burnout" as a lead-in term for "occupational stress", i.e., it treats the two terms as synonyms. Should they be considered synonyms? How should we decide this question? By considering these two terms as synonyms, the thesaurus makes it impossible to discriminate them (of course, one may search titles or other free text options, but then the thesaurus makes itself superfluous). The literature about these subjects is very large in 2019 alone, 426 titles in the PsycINFO database contained the word "burnout", while only 24 documents used the term "occupational stress" in the title, a total of 981 documents were found using this term in "all text" and the union set of the two terms in "all text" was 23,679 documents in 2019 alone. Whether or not these two terms should be considered synonyms depends on whether two distinct literatures may be distinguished so that searchers would find such a distinction useful when exploring this enormous literature. This example shows what is meant by saying that homonymity is not an inherent quality but a human choice, the usefulness of which may depend on the size and characteristics of the literature represented in a given database and thus is relative.

Two words are not synonyms but may be considered synonyms, although synonyms often are published in special dictionaries and seemingly appear to be independent of contexts.

The conclusion concerning words or symbolic structures is that they often are parts of KOS, but not necessarily. We have also claimed that universals (as defined by Herre) cannot form part of KOS, as there is no way to include the corresponding universals as something in addition to concepts. This will be explored in more detail in the following Section.

5.4. Pragmatic Realism

So far, we have encountered two different kinds of realism: "Smithian realism" and Herre's "integrative realism" and we saw that "Smithian realism" was an attempt to avoid forms of psychologism and idealism. Herre refers to Searle [92] and his claim of realism being based on the construct "common intentionality" which is a result of communication and society. However, as Kuhn [84] demonstrated, it is not enough that a given group of people connects the same meaning to something (e.g., define the term "planet" in the same way) because after a paradigm shift, the term is defined in a new way, and the former view is now understood as subjective. As we cannot exclude the possibility of yet another paradigm shift, we cannot guarantee that the new understanding is realist, as opposed

to a collective subjectivism. Therefore, universals cannot be represented, only concepts (and symbols).

How can this view be called "realism" or "materialism" as opposed to "subjectivism" and "idealism"? It is certainly not a claim that there exist specific scientific methods that can guarantee objective representations in KOS (if this was the case, we should have defended universals rather that concepts as the units). Pragmatic realism is the view that paradigms overall develop toward better representations of reality, that science makes progress (a view conflicting with Kuhn's) and that good science is more than observation and logic: it is also about values and social interests. Research reflects reality and social interests at the same time, as argued, for example, by [100], and the better this is considered, the more realist research can be. The realism of science is not given by a well-defined methodology, but by self-correcting mechanisms and by the role of pragmatic factors in theory-acceptance. While Kuhn [101] (p. 263) emphasizes how our ontologies are implied by our theories and paradigms, he nevertheless emphasizes that we cannot freely invent arbitrary structures: "Though different solutions have been received as valid at different times, nature cannot be forced into an arbitrary set of conceptual boxes. On the contrary" [. . .] "the history of the developed sciences shows that nature will not indefinitely be confined in any set which scientists have constructed so far". The world provides resistance to our conceptualizations in the form of anomalies, that is, situations in which it becomes clear that something is wrong with the structures given to the world by our concepts. In this way, Kuhn's view may be interpreted as (pragmatic) realist.

6. Conclusions

When people seek documents, information, or knowledge to answer or clarify questions, they should be given the knowledge that corresponds to our best scientific or scholarly theories and findings. Processes and systems for providing such knowledge developed in IR and KO should provide "relevant" documents to users. There has, however been a problematic tendency in information science to understand "relevance" as either based in user studies and cognitive science (in "the cognitive paradigm") or in technological issues (in "the systems view"). The neglected third position is to base relevance in subject knowledge and, ultimately, in the philosophy of science (cf., [11]).

Because Kuhn [84] refused the positivist idea of "incrementalism" (that there is a continuous accumulation of an ever-increasing stock of truths), the identification of the most qualified and relevant documents is difficult. The literature (or "the universe of recorded knowledge") in which we perform IR cannot just be considered bits of true knowledge in which each bit is as good as any other. On the contrary, the literature must be understood as a mixture of different voices, some of which conflict with each other. The relevance of a given search set of documents (or a given relevance ranking of documents) is therefore a hypothesis, the answer to which concerns the conflict between different paradigms in the subject area. This is step deeper than the problem raised by Swanson [16] that any search function forever must remain a conjecture, because it makes the adequacy of search functions—and thereby search engines and knowledge organization systems—depending on paradigms in the subject area.

Based on this insight, [12] (this conclusion is only in the online version 1.4) suggested that the domain analytic approach to classification and KO can be summarized in this way:

- Go to a given domain,
- Look at how it is classified according to contemporary knowledge (including different views),
- Discuss the basis, the epistemological assumptions and which interests are served by proposed classifications,
- Suggest a motivated classification.

Research in ontologies as KOS is today the field, in which subject specialists and philosophers are working together with computer and information scientists. As has been demonstrated in the present article, conflicting philosophies are also at play in this domain,

and we have demonstrated the need of an approach related to Kuhn's theory of paradigms in addition to pragmatic and critical theories acknowledging the role of goals, interests, and consequences in knowledge.

The article has thus provided the arguments and conceptualizations put forward in the introduction for rethinking IR and KO to provide documents and knowledge that are in accordance with our best substantiated knowledge claims.

Funding: This research received no external funding.

Acknowledgments: The author thanks the three anonymous reviewers for fine reviews, two of which increased the quality of the article.

Conflicts of Interest: The author declares no conflict of interest.

References

1. Anderson, J.D.; José, P.-C. *Information Retrieval Design: Principles and Options for Infor-mation Description, Organization, Display and Access in Information Retrieval Databases, Digital Libraries, Catalogs, and Indexes*; Ometeca Institute: St. Petersburg, FL, USA, 2005.
2. Salaba, A. Knowledge Organization Requirements in LIS Graduate Programs. *Knowl. Organ. Interf.* **2020**, *17*, 384–393. [CrossRef]
3. Salton, G. Letter to the Editor. A New Horizon for Information Science? *J. Am. J. Inf. Sci.* **1996**, *47*, 333. [CrossRef]
4. Hjørland, B. Is classification necessary after Google? *J. Doc.* **2012**, *68*, 299–317. [CrossRef]
5. Bliss, H.E. *The Organization of Knowledge in Libraries and the Subject Approach to Books*; Henry Holt: New York, NY, USA, 1933.
6. Richards, R.A. *Biological Classification: A Philosophical Introduction*; Cambridge University Press: Cambridge, UK, 2016.
7. Leydesdorff, L. Can scientific journals be classified in terms of aggregated journal-journal citation relations using the Journal Citation Reports? *J. Am. Soc. Inf. Sci. Technol.* **2006**, *57*, 601–613. [CrossRef]
8. Hjørland, B. Political Versus Apolitical Epistemologies in Knowledge Organization. *Knowl. Organ.* **2020**, *47*, 461–485. [CrossRef]
9. Hjørland, B. User-based and Cognitive Approaches to Knowledge Organization: A Theoretical Analysis of the Research Literature. *KO Knowl. Organ.* **2013**, *40*, 11–27. [CrossRef]
10. Rafferty, P. Tagging. *Knowl. Organ.* **2018**, *45*, 500–516. [CrossRef]
11. Hjørland, B. The foundation of the concept of relevance. *J. Am. Soc. Inf. Sci. Technol.* **2010**, *61*, 217–237. [CrossRef]
12. Hjørland, B. Domain Analysis. *Knowl. Organ.* **2017**, *44*, 436–464. [CrossRef]
13. Warner, J. *Human Information Retrieval*; The MIT Press: Cambridge, MA, USA, 2010.
14. Hjørland, B. Classical Databases and Knowledge Organization: A Case for Boolean Retrieval and Human Decision-making During Searches. *J. Assoc. Inf. Sci. Technol.* **2015**, *66*, 1559–1575. [CrossRef]
15. Plato. 380 B.C.E. Meno. Translated by Benjamin Jowett. Available online: https://www.gutenberg.org/files/1643/1643-h/1643-h.htm#link2H_4_0003 (accessed on 17 March 2021).
16. Swanson, D.R. Undiscovered Public Knowledge. *Libr. Q.* **1986**, *56*, 103–118. [CrossRef]
17. Hjørland, B. Citation analysis: A social and dynamic approach to knowledge organization. *Inf. Process. Manag.* **2013**, *49*, 1313–1325. [CrossRef]
18. Araújo, P.C.D.; Castanha, R.C.G.; Hjørland, B. *Citation Indexing and Indexes. Knowl. Organ.* **2021**, *48*, 58–87.
19. Hjørland, B.; Nielsen, L.K. Subject Access Points in Electronic Retrieval. *Annu. Rev. Inf. Sci. Technol.* **2001**, *35*, 249–298.
20. Turtle, H.R.; Croft, W.B. A Comparison of Text Retrieval Models. *Comput. J.* **1992**, *35*, 279–290. [CrossRef]
21. Fiorini, N.; Canese, K.; Starchenko, G.; Kireev, E.; Kim, W.; Miller, V.; Osipov, M.; Kholodov, M.; Ismagilov, R.; Mohan, S.; et al. Best Match: New relevance search for PubMed. *PLoS Biol.* **2018**, *16*, e2005343. [CrossRef]
22. Sampson, M.; Nama, N.; O'Hearn, K.; Murto, K.; Nasr, A.; Katz, S.L.; Macartney, G.; Momoli, F.; McNally, J.D. Creating enriched training sets of eligible studies for large systematic reviews: The utility of PubMed's Best Match algorithm. *Int. J. Technol. Assess. Health Care* **2021**, *37*, 1–6. [CrossRef]
23. Harter, S.P. *Online Information Retrieval: Concepts, Principles, and Techniques*; Academic Press: New York, NY, USA, 1986.
24. Frei, H.-P.; Qiu, Y. Effectiveness of Weighted Searching in an Operational IR Environment. In *Information Retrieval '93, von der Modellierung zur Anwendung; Proceedings der 1. Tagung Information Retrieval '93*; Universität Verlag Konstanz: Konstanz, Germany, 1993; pp. 41–54. Available online: http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.56.7021&rep=rep1&type=pdf (accessed on 17 March 2021).
25. Baeza-Yates, R.; Ribeiro-Neto, B. *Modern Information Retrieval: The Concepts and Technology behind Search*, 2nd ed.; Addison Wesley: New York, NY, USA, 2011.
26. Manning, C.D.; Raghavan, P.; Schütze, H. *An Introduction to Information Retrieval*; Cambridge University Press: Cambridge, UK, 2009. Available online: http://nlp.stanford.edu/IR-book/pdf/irbookonlinereading.pdf (accessed on 17 March 2021).
27. Roelleke, T. Information Retrieval Models: Foundations and Relationships. *Synth. Lect. Inf. Concepts Retr. Serv.* **2013**, *5*, 1–163. [CrossRef]
28. Salton, G.; Wong, A.; Yang, C.-S. A vector space model for automatic indexing. *Commun. ACM* **1975**, *18*, 613–620. [CrossRef]
29. Robertson, S.E.; Jones, K.S. Relevance weighting of search terms. *J. Am. Soc. Inf. Sci.* **1976**, *27*, 129–146. [CrossRef]

30. Robertson, S.E. The State of Information Retrieval. ISKO-UK. 2008. Available online: https://web.archive.org/web/2019051212 3726/http://event-archive.iskouk.org/sites/default/files/robertson.pdf (accessed on 17 March 2021).

31. Hjørland, B. Epistemology and the socio-cognitive perspective in information science. *J. Am. Soc. Inf. Sci. Technol.* **2002**, *53*, 257–270. [CrossRef]

32. Hjørland, B. Subject (of Documents). *Knowl. Organ.* **2017**, *44*, 55–64. [CrossRef]

33. Paris, L.A.H.; Tibbo, H.R. Freestyle vs. Boolean: A Comparison of Partial and Exact Match Retrieval Systems. *Inf. Process. Manag.* **1998**, *34*, 175–190. [CrossRef]

34. Robertson, S.E.; Thompson, C.L. Weighted Searching: The CIRT Experiment. In *Lnformatics 10: Prospects for Intelligent Retrieval, Proceedings of the Conference Jointly Sponsored by Aslib, the Aslib Informatics Group and the Information Retrieval Specialist Group of the British Computer Society, King's College, Cambridge, UK, 21–23 March 1989*; Karen, S.J., Ed.; Aslib: London, UK, 1990; pp. 153–165.

35. Belkin, N.J.; Croft, W.C. Retrieval Techniques. *Annu. Rev. Inf. Sci. Technol.* **1987**, *22*, 109–145.

36. Robertson, S.E. Salton Award Lecture on theoretical argument in information retrieval. *ACM Sigir Forum* **2000**, *34*, 1–10. [CrossRef]

37. Dragusin, R.; Petcu, P.; Lioma, C.; Larsen, B.; Jørgensen, H.L.; Cox, I.J.; Hansen, L.K.; Ingwersen, P.; Winther, O. FindZebra: A search engine for rare diseases. *Int. J. Med Inform.* **2013**, *82*, 528–538. [CrossRef]

38. Dragusin, R.; Petcu, P.; Lioma, C.; Larsen, B.; Jørgensen, H.L.; Cox, I.J.; Hansen, L.K.; Ingwersen, P.; Winther, O. Specialized tools are needed when searching the web for rare disease diagnoses. *Rare Dis. (AustinTex.)* **2013**, *1*, e25001. [CrossRef]

39. Acharya, A.; Verstak, A.; Suzuki, H.; Henderson, S.; Iakhiaev, M.; Lin, C.C.Y.; Shetty, N. Rise of the Rest: The Growing Impact of Non-Elite Journals. *arXiv* **2014**. Available online: http://arxiv.org/pdf/1410.2217v1.pdf (accessed on 17 March 2021).

40. Picard, C.-F.; Durocher, S.; Gendron, Y. Desingularization and Dequalification: A Foray Into Ranking Production and Utilization Processes. *Eur. Acc. Rev.* **2019**, *28*, 737–765. [CrossRef]

41. Hjørland, B. Evidence-based practice: An analysis based on the philosophy of science. *J. Am. Soc. Inf. Sci. Technol.* **2011**, *62*, 1301–1310. [CrossRef]

42. Lardera, M.; Hjørland, B. Keyword. In *ISKO Encyclopedia of Knowledge Organization*; Hjørland, B., Gnoli, C., Eds.; International Organization of Knowledge Organization (ISKO): Toronto, ON, Canada, 2020. Available online: https://www.isko.org/cyclo/keyword (accessed on 17 March 2021).

43. Mazzocchi, F. Knowledge organization system (KOS). *Knowl. Organ.* **2018**, *45*, 54–78. [CrossRef]

44. Olensky, M. Semantic Interoperability in Europeana: An Examination of CIDOC CRM in Digital Cultural Heritage Documentation. *Bull. IEEE Tech. Comm. Digit. Libr.* **2010**, *6*. Available online: https://web.archive.org/web/20130620181231/https://www.ieee-tcdl.org/Bulletin/v6n2/Olensky/olensky.html (accessed on 17 March 2021).

45. Blake, J. Some Issues in the Classification of Zoology. *Knowl. Organ.* **2011**, *38*, 463–472.

46. ISO 25964-1: 2011 (E). *Information and Documentation—Thesauri and Interoperability with Other Vocabularies—Part 1: Thesauri for Information Retrieval*; International Organization for Standardization: Geneva, Switzerland, 2011.

47. UNESCO Thesaurus. Available online: http://vocabularies.unesco.org/browser/thesaurus/en/ (accessed on 17 March 2021).

48. Wächter, T.; Alexopoulou, D.; Dietze, H.; Hakenberg, J.; Schroeder, M. *Searching Biomedical Literature with Anatomy Ontologies. Anatomy Ontologies for Bioinformatics*; Springer: London, UK, 2008; Volume 6, pp. 177–194.

49. The Foundational Model of Anatomy ontology (FMA). Available online: http://sig.biostr.washington.edu/projects/fm/AboutFM.html (accessed on 17 March 2021).

50. Aitchison, J.; Gilchrist, A.; Bawden, D. *Thesaurus Construction and Use: A Practical Manual*, 4th ed.; Aslib: London, UK, 2000.

51. Svenonius, E. Definitional Approaches in the Design of Classification and Thesauri and Their Implications for Retrieval and Automatic Classification. In *Knowledge Organization for Information Retrieval*; McIlwaine, I.C., Ed.; International Federation for Information and Documentation: The Hague, The Netherlands, 1997; pp. 12–16.

52. Hudson, M. Preparing Terminological Definitions for Indexing and Retrieval Thesauri: A Model. *Adv. Knowl. Organ.* **1996**, *5*, 363–369.

53. Gruber, T.R. A Translation Approach to Portable Ontology Specifications. *Knowl. Acquis.* **1993**, *5*, 199–220. [CrossRef]

54. Genesereth, M.R.; Nilsson, N.J. *Logical Foundations of Artificial Intelligence*; Morgan Kaufmann: Los Altos, CA, USA, 1987.

55. Colomb, R.M. *Ontology and the Semantic Web*; IOS Press: Amsterdam, The Netherlands, 2007.

56. Soergel, D.; Lauser, B.; Liang, A.; Fisseha, F.; Keizer, J.; Katz, S. Reengineering Thesauri for New Application: The AGROVOC Example. *J. Digit. Inf.* **2004**, *4*. Available online: https://journals.tdl.org/jodi/index.php/jodi/article/view/112/111 (accessed on 17 March 2021).

57. Ali, N.M.; Khan, H.A.; Amy, Y.; Then, H.; Ching, C.V.; Gaur, M.; Dhillon, S.K. Fish Ontology framework for taxonomy-based fish recognition. *PeerJ* **2017**, *5*, e3811. [CrossRef] [PubMed]

58. Rosse, C.; Mejino, J.L.V. The Foundational Model of Anatomy Ontology. In *Anatomy Ontologies for Bioinformatics: Principles and Practice*; Albert, B., Duncan, D., Richard, B., Eds.; Springer: London, UK, 2008; pp. 59–117.

59. Smith, B.; Ceusters, W.; Klagges, B.; Köhler, J.; Kumar, A.; Lomax, J.; Mungall, C.; Neuhaus, F.; Rector, A.L.; Rosse, C. Relations in biomedical ontologies. *Genome Biol.* **2005**, *6*, R46. [CrossRef] [PubMed]

60. Blunauer, A.; Pellegrini, T. Semantic Web und Semantische Technologien: Zentrale Begriffe und Unterscheidungen. In *Tassilo Pellegrini and Andreas Blumauer*; Springer: Berlin, Germany, 2006; pp. 9–25.

61. Obrst, L. *Ontological Architectures. Theory and Applications of Ontology: Computer Applications*; Springer: Dordrecht, The Netherlands, 2010; pp. 27–66.

62. Garshol, L.M. Metadata? Thesauri? Taxonomies? Topic Maps! Making Sense of It All. *J. Inform. Sci.* **2004**, *30*, 378–391. [CrossRef]
63. Aitchison, J. A Classification as a Source for a Thesaurus: The Bibliographic Classification of H. E. Bliss as a source of Thesaurus Terms and Structure. *J. Doc.* **1986**, *42*, 160–181. [CrossRef]
64. Temmerman, R. Questioning the univocity ideal. The difference between socio-cognitive Terminology and traditional Terminology. *Hermes J. Lang. Commun. Bus.* **1997**, *18*, 51–90. Available online: https://tidsskrift.dk/her/article/view/25412/22333 (accessed on 17 March 2021). [CrossRef]
65. Van Deemter, K. *Not Exactly: In Praise of Vagueness*; Oxford University Press: Oxford, UK, 2010.
66. Svenonius, E. The Epistemological Foundations of Knowledge Representations. *Libr. Trends* **2004**, *52*, 571–587.
67. Hjørland, B. Are Relations in Thesauri 'Context-Free, Definitional, and True in all Possible Worlds'? *J. Assoc. Inf. Sci. Technol.* **2015**, *66*, 1367–1373. [CrossRef]
68. Fidel, R. Searchers' Selection of Search Keys: II. Controlled Vocabulary or Free-Text Searching. *J Am Soc Inform Sci.* **1991**, *42*, 501–514. [CrossRef]
69. Dahlberg, I. Brief Communication: Concepts and Terms—ISKO's Major Challenge. *Knowl. Organ.* **2009**, *36*, 169–177. [CrossRef]
70. Hjørland, B. Fundamentals of Knowledge Organization. *Knowl. Organ.* **2003**, *30*, 87–111.
71. Sowa, J.F. *Conceptual Structures: Information Processing in Mind and Machine*; Addison-Wesley: Reading, MA, USA, 1984.
72. Patel, A.; Jain, S.; Shandilya, S.K. Data of Semantic Web as Unit of Knowledge. *J. Web Eng.* **2019**, *17*, 647–674. [CrossRef]
73. Machado, L.; Simões, G.; Gnoli, C.; Souza, R. Can an Ontologically-Oriented KO Do Without Concepts? *Knowl. Organ. Interf.* **2020**, *17*, 502–506. [CrossRef]
74. Herre, H. General Formal Ontology (GFO): A Foundational Ontology for Conceptual Modelling. In *Theory and Applications of Ontology: Computer Applications*; Roberto, P., Michael, H., Achilles, K., Eds.; Springer: Dordrecht, The Netherlands, 2010; pp. 297–345. [CrossRef]
75. Smith, B. Beyond Concepts: Ontology as Reality Representation. In *Proceedings of the FOIS 2004. International Conference on Formal Ontology and Information Systems, Turin, Italy, 4–6 November 2004*; Achille, C.V., Laure, V., Eds.; IOS Press: Amsterdam, The Netherlands, 2004. Available online: https://www.researchgate.net/publication/244107491_Beyond_Concepts_Ontology_as_Reality_Representation (accessed on 7 February 2021).
76. Smith, B.; Ceusters, W. Ontology as the Core Discipline of Biomedical Informatics. In *Compu-ting, Philosophy, and Cognitive Science: The Nexus and the Liminal*; Susan, A.J.S., Gordana, D.C., Eds.; Cambridge Scholars Press: Newcastle, UK, 2007; pp. 104–122.
77. Arp, R.; Smith, B.; Spear, A.D. *Building Ontologies with Basic Formal Ontology*; The MIT Press: Cambridge, MA, USA, 2015.
78. Leclercq, H. Europe: Term for Many Concepts. *Int. Classif.* **1978**, *5*, 156–162. [CrossRef]
79. Smith, B.; Ceusters, W. Ontological realism: A methodology for coordinated evolution of scientific ontologies. *Appl. Ontol.* **2010**, *5*, 139–188. [CrossRef]
80. Leonelli, S. *Data-Centric Biology: A Philosophical Study*; University of Chicago Press: Chicago, IL, USA, 2016.
81. Ogden, C.K.; Richards, I.A. *The Meaning of Meaning: A Study of the Influence of Language Upon Thought and of the Science of Symbolism*; Routledge & Kegan Paul: London, UK, 1923.
82. Sowa, J.F. Ontology, Metadata, and Semiotics. In *Lecture Notes in Computer Science*; Springer: Berlin, Germany, 2000; Volume 1867, pp. 55–81.
83. Hjelmslev, H. *Omkring Sprogteoriens Grundlæggelse*; B. Lunos bogtrykkeri: Copenhagen, Denmark, 1943.
84. Kuhn, T.S. *The Structure of Scientific Revolutions*; University of Chicago Press: Chicago, IL, USA, 1962.
85. Thagard, P. Conceptual Change. In *Encyclopedia of Cognitive Science*; Nadel, L., Ed.; Macmillan: London, UK, 2003; Volume 1, pp. 666–670. Available online: http://cogsci.uwaterloo.ca/Articles/conc.change.pdf (accessed on 17 February 2021).
86. Del Hoyo, J.; Elliott, A.; Sargatal, J.V. (Eds.) *Handbook of the Birds of the World*; Lynx Edicions: Barcelona, Spain, 1997; Volume 1.
87. Fjeldså, J. Avian Classification in Flux. In *Handbook of the Birds of the World*; Lynx Edicions: Barcelona, Spain, 2013; Special Volume 17, pp. 77–146.
88. Mayr, E.; Bock, W.J. Provisional Classifications v Standard Avian Sequences: Heuristics and Communication in Ornithology. *IBIS* **1994**, *136*, 12–18. [CrossRef]
89. Sibley, C.; Ahlquist, J.E. *Phylogeny and Classification of Birds: A Study in Molecular Evolution*; Yale University Press: New Haven, CT, USA, 1990.
90. Hjørland, B. Concept Theory. *J. Am. Soc. Inf. Sci. Technol.* **2009**, *60*, 1519–1536. [CrossRef]
91. Biagetti, M.T. Ontologies (as Knowledge Organization Systems). *ISKO Encycl. Knowl. Organ.* **2020**. Available online: https://www.isko.org/cyclo/ontologies (accessed on 17 March 2021).
92. Searle, J. *The Construction of Social Reality*; Free Press: New York, NY, USA, 1995.
93. Wardy, R. Categories. In *Routledge Encyclopedia of Philosophy*; Edward, C., Ed.; Routledge: London, UK, 1998; Volume 1–10.
94. Thomasson, A. Categories. In *The Stanford Encyclopedia of Philosophy*; Edward, N.Z., Ed.; 2019. Available online: https://plato.stanford.edu/archives/sum2019/entries/categories/ (accessed on 17 March 2021).
95. Moss, W.R. Categories and Relations: Origins of Two Classification Theories. *Am. Doc.* **1964**, *15*, 296–301. [CrossRef]
96. Furner, J. Type–Token Theory and Bibliometrics. In *Theories of Informetrics and Scholarly Communication*; De Gruyter Saur: Berlin, Germany, 2016; pp. 119–147.
97. Fellbaum, C. WordNet. In *Theory and Applications of Ontology: Computer Applications*; Springer: Dordrecht, The Netherlands, 2010; pp. 231–243.

98. Caputo, J.D. *Hermeneutics: Facts and Interpretation in the Age of Information*; Penguin: London, UK, 2018.
99. Soergel, D. WordNet [Book Review]. *D-Lib Mag.* **1998**, *4*, 1–7. Available online: http://www.dlib.org/dlib/october98/10 bookreview.html (accessed on 17 March 2021).
100. Barnes, B.; Bloor, D.; Henry, J. *Scientific Knowledge: A Sociological Analysis*; The University of Chicago Press: Chicago, IL, USA, 1996.
101. Kuhn, T.S. Reflections on My Critics. In *Criticism and the Growth of Knowledge*; Lakatos, I., Musgrave, A., Eds.; Cambridge University Press: Cambridge, UK, 1970; pp. 231–278.