Wiesław Babik*

Keywords as Linguistic Tools
for Information and Knowledge Organization

Abstract: Information and knowledge organization by keywords is presently effected in information and retrieval systems – not only in various taxonomic structures, but also in new structures of a folksonomic and ontological nature as well as mind maps.

The need for a fresh look at keyword results from the new possibilities of document and information indexing, especially of social (collaborative) indexing, among others. This will require new research and scholarly reflection on information processing by both computer systems and the human mind. And this will, in turn, create the need to look at both information and knowledge organization and keywords in terms of the network organization of information and knowledge. This topic is current, owing to the application of keywords in the WWW system, subject indexes, OPACs, full-text systems, and digital libraries. The considerations are concluded by an attempt to answer the question contained in the title relating to the roles of keywords in information and knowledge organization for current information systems. This study has shown that, by using keywords, we essentially deal with a natural language in a meta-information function, not only with an artificial keyword language; "-nomic" systems based on hierarchy are usually strongly structured, while the systems that use keywords are hardly structured. Information and knowledge organization by keywords is frequently based on the indexing practice, which means that it is carried out bottom up (with classification being a reverse process).

Keywords: keywords, tags, keyword indexing, tag cloud, knowledge organization, information organization, folksonomies, ontologies, taxonomies, mind maps

*   Jagiellonian University, Institute of Information and Library Science, Krakow, Poland

1.   Introduction

The goal and thread of my considerations consist of a theoretical reflection on the scope and current possibilities of using keywords (more and more frequently called "tags") for the information and knowledge representation and organization. This issue will be presented by a comparative method and evaluation methodology[1]. The question of how far keywords are meta-data worthy of indexing will be answered. In the same context, I will also discuss the issue of the identity of keywords (in comparison to classical and non-classical keyword languages) as well as their relationships with a natural language. Methodological and theoretical backgrounds will be presented as well a recommendation for an indexing practice and its application in indexing and retrieval systems.

Recently, keyword[2] roles in information and knowledge organization have been discussed by the team Ramatollah Fattahi, Mehri Parirokh, Mahammd H. Dayyani, and Abdolrasoul Khosravi [9], and in Poland by Wiesław Babik [1][3] , Anna Górska [10], Piotr Matlak [17], and Wiesław Babik [2].

2.   Information and Knowledge Organization: Definitions and Aspects

Theoretical foundations of information and knowledge organization with their historical, sociological, and organization backgrounds have been recently presented by H. Peter Ohly [19, 20], Birger Hjørland [12], Richard P. Smiraglia [22, 23], Manfred Hauer [11], Marjorie M.K. Hlava [16], and Ingetraut Dahlberg [7].

Ingetraut Dahlberg defines "Knowledge Organization" as "[...] the science of structuring and systematically arranging of knowledge units (concepts) according to their inherent knowledge elements (characteristics) and the application of concepts and classes of concepts ordered by this way for the assignment of the worthwhile contents of referents (objects/subjects) of all kinds" [6].

Birger Hjørland [12] provided six theoretical approaches to knowledge organization:
  – traditional approaches;
  – facet analytic approaches;

---

[1] This paper attempts to show the new viewpoint on problems presented in: W. Babik, *Keywords as Linguistic Tools in Information and Knowledge Organization*, [in:] Babik W., Ohly H.P., Weber K. (eds.), *Theorie, Semantik und Organization von Wissen*, Ergon Verlag, Würzburg 2017, pp. 306–313.

[2] Frequent use of the term of "keyword" is confusing. It appears to refer to controlled vocabulary terms, while in English, a "keyword" most-often refers to either free text or terms assigned by authors.

[3] The study contains terminological arrangements, formulates premises about keywords, and presents the results of keyword analysis in various contexts. The keyword language, as an abstract and formal structure, was subjected to a detailed functional and structural description. Theoretical considerations were enriched by the results of the author's research on the keyword vocabularies of traditional systems and their applications in the Internet. These considerations put the current conceptions of keywords and keyword languages into order.

- information retrieval tradition;
- user-oriented and cognitive views;
- bibliometric approaches;
- domain analytic approach:
  (http://en.wikipedia.org/wiki/Knowledge_organization).

Based on his analysis the above-mentioned author concluded that knowledge organization and information organization were hardly different from each other and were, in fact, quasi-synonymous; therefore, the phrase *information and knowledge organization* is used in the title of this paper. The considerations voiced in my article belong to that stream of consideration that B. Hjørland called a "user-based and cognitive approach to knowledge organization". My considerations constituted a specific continuation of Hjørland's studies conducted in his recently published articles on knowledge organization [13, 14].

3. Theoretical Background

Present-day thinking becomes more and more free association-based, metaphorical, and multi-aspectual rather than strict, logical, and linear. "Instead of ordering knowledge under various clear-cut classes, the information society offers to use cascades of non-contextual signs, inter-connected in a more or less accidental manner" [8]. This statement by Thomas H. Eriksen perfectly describes the keywords currently being applied in information and knowledge organization.

Keywords only appeared in the early 1950's, although their origin can be traced to a much-earlier time; e.g., in subject head words (Middle Ages) or the uniterm language. Keywords appeared as a result of the need to search for new document indexing methods, taking into account the possibilities introduced by computers (computer technologies) and based on the bottom-up trend in the approach to indexing processes. Presently, the same is achieved in the form of folksonomy, based on the possibilities of the human mind, using current developments of neural science (neuroscience) and the collective intelligence capabilities. Since keywords are included among linguistic tools called indexing and retrieval languages, it is justified to look at them and treat them – similar to other indexing and retrieval languages – as information and knowledge organization tools.

Keywords are widely used in the WWW system, subject indexes, OPAC's, full-text systems, and digital libraries, for example. Undoubtedly, these are various examples of using keywords, taking into account the diverse functions they offer for present-day information systems.

If we perceive keywords from a linguistic viewpoint, their network recognition within a network information and knowledge organization will be a proper plane of consideration. What is indispensable here, in my opinion, is the differentiation between three terms: system (Fig. 1), network (Fig. 2), and hybrid structure (Fig. 3).

Fig. 1. Two hierarchical structures (systems)

The system is clearly a hierarchical structure and is characterized by the fact that each of its elements (except the first) is directly hard asymmetric, dependent only on one or more than one other element.
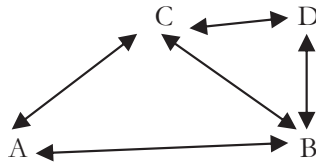

Fig. 2. Network

And it is characteristic for a network structure that its particular elements can be soft free symmetric, dependent on more than one element.
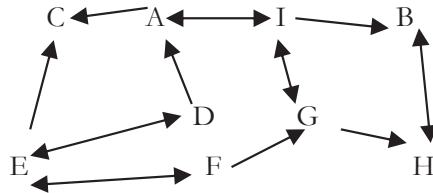

Fig. 3. Hybrid structure

Of course, an intermediate form can exist: a hybrid structure. And there are more than one of these structures. Accounting for this, we distinguish knowledge organization models: a systemic and network one, next to the hybrid models. Keywords generally and actually constitute a network, although they can occur locally within various subsystems that can have the nature of a full system. It is similar to the case of the WWW system.

A classical keyword language is a typical network of keywords as lexical units of that language. At this point, I disregard non-classical variations of that language that are not classical networks but rather specific types of hybrids, although they

also correspond to the human mind and natural language properties. Network thinking of the present-day society or scientific community is not equivalent to systemic thinking. Under the influence of the Internet, people are unable to think systemically (that is, hierarchically) or with relational cohesion; at most, they can do it with some difficulty, especially in the case of young people. People rather think in terms of a network. However, logical thinking is based on a systemic approach (unlike network thinking, which is based on loose associations and impressions).

Therefore, keywords belong to the network information and knowledge organization where one uses keywords, based primarily on the natural language competence and assuming the paradigmatic system of that language. A network approach to keywords is currently realized first of all in the various indexing and retrieval languages of a folksonomic nature. These structures are bottom-up structures, just as folksonomies and mind maps; but, in that language's class, taxonomic (top-down) structures exist, simply as classifications and ontologies.

Do keywords constitute indexing and retrieval languages? Or are they maybe a natural language? [1, 4]. I support the recognition that classical keywords constitute the natural language in the retrieval function, using current computer technology capabilities. But this is gradually changing. Is the keyword language evolving from an uncontrolled form [1] to a controlled one [10], owing to new computer technology capabilities? The present stage of this evolution is undoubtedly represented by tags. I think that the term "keyword" will disappear in the future. It will instead be replaced by the term "tag." Even now, it is easy to notice the coexistence of both terms (although with a clear tendency to abandon the former). Thus, tags will remain and consume the present keywords.

How are keywords different as linguistic tools of information and knowledge organization from other things? Regardless of variation, keywords are *-nomic* systems that can perform various functions or roles. In fact, they constitute the natural language in different variations. This language, however, offers a different method of information and knowledge organization than classifications (in fact, the latter class of indexing and retrieval languages is also diverse). These are usually artificial systems (in the sense of origin), and they have either digital or alphanumeric notations. Less often, they use a "natural" notation. Classical keyword languages are mostly based on bottom-up indexing, in contrast to typical classifications (or classification languages) that are based on top-down indexing.

4. Network Methodology: A New View on Keywords

Until recently, we were looking at the world through systemic glasses; however, presently influenced by the Internet, we are looking primarily through network glasses [18]. Separate networks usually have complex structures, composed of nodes

and connecting relationships. The network approach undoubtedly has holistic ambitions consisting of the holistic treatment of phenomena. Using Manuel Castells' term of a "network society", we rather point out the method or form of something (or communication), not the contents. Keywords usually play the role of node names that are somehow connected to each other by the relationships expressed in the given natural language. The lexical system of a keyword language is more of a network than a system. Keywords display such features, being elements of network information and knowledge organization. This specific kind of freedom allows us to organize knowledge with keywords almost in an arbitrary way (depending on man's or a system's needs and capabilities) as well as reveal the creative abilities of man as an organizer. Consequently, we can uncover new arrangements and new knowledge (data mining, knowledge discovery). Owing to the arbitrariness offered by the network, we can create bottom-up structures and discover various structures without being limited to strict structures related to the system that are planned in advance (e.g., to classification).

Regardless of their types, keywords play essential roles in the presently networked information world (although keywords are not unique in this respect). They offer possibilities that are not available in typical classifications. Consequently, keywords fill in a certain gap in this area. It is true that classifications start with general ideas and reach specific ones. In the case of keywords, this is not so obvious. Keywords are excellent tools for representing the world and knowledge of unclear boundaries and the so-called disintegrated world. Some keyword languages also provide detailing possibilities. It is not a specialist who presently decides this, but rather ordinary folks. Anyway, the differences existing between experts and ordinary people (as to the choice of keywords) are diminishing. Thus, keywords are reaching favorable areas of use. Consequently, we observe changes in the indexing culture (and also regarding keyword indexing). In this area, it is not experts (authors' keywords) or librarians who decide the choice of keywords, but rather laymen. Experts may care about the unification of keywords the most, depending on their needs and system capabilities. This process leads to changes in keyword forms. More and more, keyword vocabularies often do not include single words/terms but rather long and complex key phrases; e.g., "book, library, and information sciences", or "information and communication models in book science". Vocabulary standardization starts to be a domain of the indexing or retrieval system, and it stops being controlled by the system user.

To what extent are the lay creations valuable? Presently, we are dealing with both the wisdom and stupidity of the mob, and nobody can answer this question explicitly. The answer can certainly depend on the type of indexing community. The community decides on the quality of the indexing process. Therefore, keywords are suitable for a "think network" of the present-day man using the Internet.

Information and knowledge organization with the use of keywords is a horizontal and flat method. It does not display any hierarchical structure nor vertical tag

collection, causing it to be rather difficult to reach information. The FaceTag method allows us to avoid this difficulty.

The method is perceived in a similar fashion to clusters. It operates with hierarchical tags. These allow us to determine hierarchical relationships and present the respective dependencies between the objects being described. A tag becomes an equivalent of a face, and a single object obtains a description with many faces. Thus, the object can be comprised holistically or completely, and from many different prospects. Folksonomy is actually meta-information (metadata), which is neither top-down nor centrally controlled but rather bottom-up controlled. Consequently, the same information can be described in many different ways at the same time, giving us the possibility of reaching it in many different ways as well. We need to acknowledge that folksonomies are essentially different from classical classifications. They do not create any hierarchy of information on the Internet, but rather evaluate tags (tag clouds) without arranging them in a hierarchy.

Since a considerable majority of keywords are terms, they are also used in ontology building [5]. They also play the role of object naming, and this is why they are used in mind maps [24].

5.   Recommendations for Indexing Processes in Retrieval Systems

People think that social categorization (folksonomy) consists of the creation of keywords by service users, not by qualified personnel; this fact can positively influence the increase of information retrieval effectiveness. To make their lives easier, people apply specific, spontaneous, non-scientific typologies, segmenting in that way a particular fragment of reality (including documentary reality). By themselves, Internet users create something like a self-controlled system, inventing the categories that they currently need. That is because John Battelle [3] defined folksonomy as "a new marking scheme not based on a precise top-down hierarchy but rather on an unordered, bottom-up solution". This is undoubtedly a valuable method of organizing and ordering Internet contents based on markers called tags. Tagging means describing the contents with any keywords, with strongly present social and public elements and not by the use of any controlled vocabulary. Marcin Roszkowski's definition [21] is very appropriate in this respect. In contrast to classification, where "each document has its place like a leaf on a branch. However, in respect to folksonomy, there is no tree on which one can hang a leaf. Leaves of the same shape are <raked> in one place, and those of a different color in a different place".

In the indexing process, folksonomies undoubtedly contribute to a reduction of the barrier between experts and regular users in the matter of keyword selection

and, consequently, in the selection of a proper document indexing language, because everyone becomes a user. Secondly, it is the task of a smart system/software to standardize keywords/tags when a user instead applies the natural language. In this way, tags become elements of the natural language.

How is it possible to evaluate the indexing process used by the mob? Undoubtedly, education and saturation with "wise men" is required for this purpose, as well as making our computers smarter (although, even now, they can resolve a number of serious problems for which intelligence is needed) [4].

It is quite possible that we will live in a hybrid world in the future. In fact, such a world exists now! We live in a "multi-" world where the network and system are intermingled in various ways. Thus, keywords are networks globally and often systems locally. Such heterogeneity of the information and knowledge organization structure corresponds to characteristic features of present-day man's thinking, both in the positive and negative sense. Keywords are perfectly included in the network paradigm of thinking about information retrieval languages as information and knowledge organization tools.

6. Conclusions

A systemic approach to keywords allows us to study them exclusively within a system structure and, thus, for the exposition of structural elements by a reduction of keyword pragmatic studies to some extent. The network approach allows us to notice keyword connections with mental processes and the dynamics of an information retrieval organization using keywords. It also allows us to recognize the needs of users (who are often under the pressure of network and other co-operating systems) as deciding factors. The semantic field of free keywords has a fuzzy nature – similar to the natural language – and the field boundaries are changeable. Using keywords, we obtain a full picture of reality being described, albeit cut into pieces and incongruent, including documentary reality and, consequently, a network organization of information and knowledge. The structures originating from such an organization do not take into account any hierarchical relationships, assuring the cohesion and precision of retrieval. They do, however, take into account associative relationships (including relatedness), assuring the completeness of information retrieval [1].

Classifications integrate and order fragmented and dispersed information and knowledge, while keywords do it in the opposite direction: they disperse and subdivide existing information and knowledge. A user of an information retrieval system can select which information or knowledge he offers or needs, either in whole or in part, and choose a proper method of information ordering and presentation as a consequence.

References

[1] Babik W.: *Słowa kluczowe* [*Keywords*]. Wydawnictwo Uniwersytetu Jagielloń-skiego, Kraków 2010.

[2] Babik W.: *Układy gniazdowe terminów i słowniki słów kluczowych w systemie informacji z zakresu etnologii i antropologii kulturowej* [*The nesting systems of terms and keyword vocabularies in the information system designed for ethnology and cultural anthropology*]. [in:] Barański J., Golonka-Czajkowska M., Niedźwiedź A. (red.), *W krainie metarefleksji*, Wydawnictwo Uniwersytetu Jagiellońskie-go, Kraków, 2015.

[3] Battelle J.: *Szukaj. Jak Google i konkurencja wywołali biznesową i kulturową rewolucję* [*The search: How Google and its rivals rewrote the rules of business and transformed our culture*]. Wydawnictwo Naukowe PWN, Warszawa 2006.

[4] Bojar B.: *Języki informacyjno-wyszukiwawcze wczoraj, dziś... czy jutro?* [*Information-Retrieval languages – Yesterday, Today and Tomorrow?*]. Zagadnienia Informacji Naukowej, nr 1(93), 2009, pp. 3–24.

[5] Bosch M.: *Ontologies, Different Reasoning Strategies, different Logics, Different Kinds of Knowledge Representation: Working Together*. Knowledge Organization, no. 3, 2006, pp. 153–159.

[6] Dahlberg I.: *Definitionen aus dem Begriffsfeld Wissensorganisation*. http://www.bonn.iz-soz.de/wiss-org/definitionen_aus_dem_begriffsfeld.htm [access: 10.02.2016].

[7] Dahlberg I.: *Wissensorganisation. Entwicklung, Aufgabe, Anwendung, Zukunft. Hrsg. von der Deutschen Sektion der Internationalen Gesellschaft für Wissensorganisation e.V. (ISKO)*. Ergon Verlag, Würzburg 2014, p. 378.

[8] Eriksen T.H.: *Tyrania chwili: szybko i wolno płynący czas w erze informacji* [*Tyranny of the Moment: Fast and Slow Time in the Information Age*]. Państwowy Instytut Wydawniczy, Warszawa 2003.

[9] Fattahi R., Parirokh M., Dayyani M.H., Khosravi A.: *The Efficacy of Google's Suggested Keywords and Phrases in Query Expansion Based on "The Least Effort Principle" and "Cognitive Load Theory"*. [in:] Neelameghan A., Raghavan K.S. (eds.), *Categories, Contexts and Relations in Knowledge Organization. Proceedings of the Twelfth International ISKO Conference 6–9 August 2012 Mysore, India*, Ergon Verlag, Würzburg 2012, pp. 378.

[10] Górska A.: *Tagowanie kontrolowane – oksymoron czy przyszłość języków informacyjno-wyszukiwawczych?* [*Controlled Tagging – Oxymoron or the Future of Information-Retrieval Languages?*]. Zagadnienia Informacji Naukowej, nr 2(100), 2012, pp. 6–17.

[11] Hauer M.: *Zur Bedeutung normierter Terminologien in Zeiten moderner Sprach- und Information-Retrieval-Technologien*. ABI Technik, Heft 1, 2013, pp. 2–6.

[12] Hjørland B.: *What is Knowledge Organization (KO)?*. Knowledge Organization, no 2/3, 2008, pp. 86–101.

[13] Hjørland B.: *Is Knowledge Organization = Information Organization?* [in:] Neelameghan A., Raghavan K.S. (eds.), *Categories, Contexts and Relations in Knowledge Organization. Proceedings of the Twelfth International ISKO Conference 6–9 August 2012 Mysore, India*, Ergon Verlag, Würzburg 2012, pp. 8–14.

[14] Hjørland B.: *User-based and Cognitive Approaches to Knowledge Organization: A Theoretical Analysis of the Research Literature*. Knowledge Organization, no. 1, 2013, pp. 11–27.

[15] Hjørland B.: *Knowledge Organization*. Knowledge Organization, vol. 43, issue 6, 2016, pp. 475–484.

[16] Hlava M.M.K.: *The Textbook: History, Theories, and Concepts of Knowledge Organization*. Morgan & Claypool, San Rafael CA 2014.

[17] Matlak P.: *Indeksowanie treści. Porównanie skuteczności metod tradycyjnych i automatycznych* [*Content Indexing. Comparizon of Efficiency of Traditional and Automatic Methods*]. Wydawnictwo SBP, Warszawa 2012.

[18] Muraszkiewicz M.: *Esej: nowy paradygmat, czyli od systemu do sieci* [*An Essay on the New Paradigm, i.e. from System to Network*]. [in:] Ed. Sosińska-Kalata B., Przastek-Samokowa M. (red.), Skrzypczak A. (współpr.), *Od informacji naukowej do technologii społeczeństwa informacyjnego* [*From Scientific Information to the Technology of Information Society*], Wydawnictwo Stowarzyszenia Bibliotekarzy Polskich, Warszawa 2005, pp. 83–86.

[19] Ohly H.P.: *Knowledge Organization Pro and Retrospective*. [in:] Arsenault C., Tennis J. (eds.), *Culture and Identity in Knowledge Organization. Proceedings of the Tenth International ISKO Conference 5–8 August 2008 Montreal, Canada*, Ergon Verlag, Würzburg 2008, pp. 2010–2015.

[20] Ohly H.P.: *Mission, Programs, and Challenges of Knowledge Organization*. [in:] Neelameghan A., Raghavan K.S. (eds.), *Categories, Contexts and Relations in Knowledge Organization. Proceedings of the Twelfth International ISKO Conference 6–9 August 2012 Mysore, India*. Ergon Verlag, Würzburg 2012, pp. 15–23.

[21] Roszkowski M.: *Folksonomia jako narzędzie społecznego tagowania* [*Folksonomy as Social Tagging Tool*]. Warsztaty Bibliotekarskie, nr 4(24), 2007, [on-line:] http://www.pedagogiczna.edu.pl/warsztat/2007/4/070404.htm [access: 10.02.2016].

[22] Smiraglia R.P.*: Universes, Dimensions, Domains, Intensions and Extensiones: Knowledge Organization for the 21st Century*. [in:] Neelameghan A., Raghavan K.S. (eds.), *Categories, Contexts and Relations in Knowledge Organization. Proceedings of the Twelfth International ISKO Conference 6–9 August 2012 Mysore, India*, Ergon Verlag, Würzburg 2012, pp. 1–7.

[23] Smiraglia R.P.: *The Elements of Knowledge Organization*. Springer, Cham 2014.

[24] Włodarczyk B.: *Mapy tematów jako system reprezentacji wiedzy* [*Topic Maps as Knowledge Representation System*]. Zagadnienia Informacji Naukowej, nr 1(99), 2012, pp. 14–34.

Słowa kluczowe
lingwistycznym narzędziem organizacji informacji i wiedzy

Streszczenie: Potrzeba nowego spojrzenia na słowa kluczowe wynika z nowych możliwości indeksowania dokumentów i informacji, zwłaszcza indeksowania społecznościowego, ale nie tylko. Wymaga nowych badań i refleksji naukowej nad przetwarzaniem informacji zarówno przez mózg człowieka, jak i w systemach komputerowych. To z kolei stwarza potrzebę spojrzenia zarówno na organizację informacji i wiedzy, jak i na słowa kluczowe w kategoriach koncepcji sieciowej organizacji informacji i wiedzy. Temat jest aktualny przede wszystkim z powodu stosowania słów kluczowych w systemie WWW, indeksach rzeczowych, katalogach OPAC, systemach pełnotekstowych czy bibliotekach cyfrowych.

Organizacja informacji i wiedzy za pomocą słów kluczowych jest obecnie urzeczywistniana w systemach informacyjno-wyszukiwawczych nie tylko w różnych strukturach o charakterze folksonomicznym, lecz także w nowych strukturach ontologicznych, taksonomicznych i mapach myśli.

Głównym wątkiem pracy jest refleksja nad zakresem i aktualnymi możliwościami wykorzystania słów kluczowych, nazywanych coraz częściej tagami, w reprezentacji oraz organizacji informacji i wiedzy. Zagadnienie to zostanie przedstawione na płaszczyźnie porównawczej. W tym kontekście zostanie poruszony również problem tożsamości używanych słów kluczowych w porównaniu z klasycznymi i nieklasycznymi językami słów kluczowych, a także ich relacje do języka naturalnego.

Rozważania kończą się próbą odpowiedzi na pytanie zawarte w tytule o rolę słów kluczowych w organizacji informacji i wiedzy we współczesnych systemach wyszukiwania informacji. W konkluzji podkreśla się, że używając słów kluczowych, w istocie mamy do czynienia przede wszystkim z językiem naturalnym w funkcji metainformacyjnej, a nie tylko ze sztucznym językiem słów kluczowych, oraz że systemy „-nomiczne" oparte na hierarchii zwykle są mocno ustrukturalizowane, zaś systemy wykorzystujące słowa kluczowe są słabo ustrukturalizowane. Organizacja informacji i wiedzy za pomocą słów kluczowych najczęściej opiera się na praktyce indeksowania, czyli dokonuje się „od dołu do góry", podczas gdy klasyfikowanie jest podejściem odwrotnym.

Słowa kluczowe: słowa kluczowe, tagi, chmura tagów, organizacja wiedzy, organizacja informacji, ontologie, taksonomie, mapy myśli