What Do we Analyze during Subject Analysis of Documents? The Concept of Aboutness in Knowledge Organization

Helena Kučerová / Institute of Information Studies and Librarianship, Philosophical Faculty, Charles University, Prague

Summary:

The study presents the results of our literary research that summarizes the approaches to the concept of "aboutness" in information science, examining the reasons for the current renewed interest in the semantics of information and knowledge-processing and the consequent efforts to develop new approaches and solutions. The paper is divided into three parts. The first part summarizes the semantic issues of content analysis. Using our own model, the concept of "aboutness" is presented, for which there is no equivalent in Czech terminology, being instead loosely expressed by other terms, such as topic, content, subject, meaning, theme, sense, and genre. The second part presents the opinions of several authorities in the field of information science on the subject-matter. The third part situates the concept of "aboutness" into FRSAD model.

Keywords: aboutness, content of documents, subject analysis, indexing, relevance, knowledge organization, FRSAD

The study is a part of NAKI DF13P01OVV013 project, *Knowledge Bases for the Discipline of Organization of Information of Knowledge* sponsored by ÚISK FF UK (Institute of Information Studies and Librarianship, Philosophical Faculty, Charles University) in Prague.

"And how are you going to go about your search for that which you do not know at all, Socrates? Which one of all the things you don't know will you make the target of your search? And even if you did come across it, how will you know that it is that which you did not know?"

Plato (427–347 BC). *Meno.*

1 Introduction

At present, there is a surge of renewed interest in the semantics of knowledge organization, accompanied by new approaches to semantic issues and their problem-solving. Interest in semantics is a logical response to the current developmental stage of communication among information resources that has moved past problems with syntax, while higher-level issues are now being considered. Attempts at innovation have been prompted by the current changes of the paradigm of knowledge organization[1] caused mainly by the massive proliferation of electronic

[1] SILVA, Armando Malheiro da a Fernanda RIBEIRO. "Documentation/Information and their Paradigms: Characterization and Importance in Research, Education, and Professional Practice." In: *Knowledge organization*. 2012, **39**(2), 111–124. ISSN 0943-7444.

documents and by the no-less massive penetration of internet technologies into daily work with information resources; that is, with documents, data and with people. The most significant semantic consequences of the changing technological paradigm of knowledge organization are the following: *1. isolation of the content of information resources 2. interconnectedness of the content of information resources 3. finer granularity of the content 4. loss of context of the content 5. dependence of access to content of information resources on computer programs.*

1. In the case of electronic documents, not only are the originally fixed ties of the content of traditional documents to their physical carrier being relaxed, but more often than not, so are the ties to the proprietary application software and the user interface. In other words, contemporary technologies make it possible to separate the processing of the content from the form of the information resources, making thus their semantics and syntax independent of each other. For example, the syntactic structure of an HTML document, marked with tags, has no impact on the content with which we fill out the individual formal parts of the document.

2. Internet technologies have made it possible to generate a network of documents interconnected via hypertext, providing the infrastructure for social networks of interlinked individuals. These networks are currently being expanded by linked data network. They operate on the principle that potentially relevant static documents are no longer stored in repositories as whole entities, but are decomposed into individual parts according to their content prior to being stored. The user then has the option to have an ad hoc dynamic document generated according to his individual request; this document is "composed" from individual, factually relevant distributed facts.[2]

3. The above-given changes to the information environment have caused that the so-far document-oriented methods of knowledge organization are being replaced with methods focusing on finer granularity of the processed content all the way to the level of individual facts (e.g.; in the format of RDF language).[3]

4. Of course, greater flexibility and the possibilities of multiple-use of detached content are not without problems; these ensue primarily from content distribution and the loss of context, which may affect the meaning of the isolated data.

5. Thus "liberated" content may only be approached via computer programs, which causes another set of problems since the programs have no intelligence of their own, being able to work with data only on the level of syntax. Supporting structures for data evaluating ("understanding") are therefore necessary for semantic processing of data. These auxiliary structures pass on the cognitive model of the world, with which thinking people work, to software applications. Pursuant to their basic building unit—concept—these structures are called conceptual data models.

In ICT, interest in semantics is showing in the form of semantic web initiative (RDF and OWL languages, ontology); in the domain of memory institutions, there are the already-mentioned conceptual data models of documents and systems of knowledge organization

[2] *Linking Open Data Cloud Diagram* by Richard Cyganiak a Anja Jentzsch illustrates the scope and content of interconnected data network, available at: http://lod-cloud.net/.

[3] "Resource Description Framework" (Rámec popisu [informačních] zdrojů). CELBOVÁ, Ludmila. RDF. In: *KTD: Česká terminologická databáze knihovnictví a informační vědy (TDKIV)* [online]. Praha: Národní knihovna ČR, 2003- [cit. 2014-04-08]. Available at: http://aleph.nkp.cz/F/?func=direct&doc_number=000000552&local_base=KTD.

(FRBR/FRAD/FRSAD model of universal bibliographic reference, SO 25964 thesaurus model, CIDOC CRM[4]/ISO 21127 model of cultural heritage information, OAIS/ISO 14721 model of open archival information system, etc.). What the two initiatives have in common is their objective to intercept and interpret the content of the communicated information resources in forms understandable not only to humans but also to computer programs. We have already mentioned here the importance of semantics for computer programs; however, according to many experts, new technology makes semantics "for humans" increasingly more important, too.

"Tools such as thesauri and classification schemata have so far been used primarily for indexing and searching. However, in the era of humongous digital archives and powerful search algorithms, construction of systems that help understanding, contextualizing and finding one's way in the mass of available sources have become key to information organization. Rather than organizing information in the vein of traditional understanding, the building of such systems resembles the creation of syllables or short excerpts, or the production of shows and exhibitions."[5]

English often uses the neologism, "aboutness," for that which is recorded and represented via content analysis of information resources. "About" is a preposition with multiple meanings. However, only its Czech equivalent "o" of the sixth declension case is relevant for our purposes as it expresses the conjunction of verbs and verbal nouns with their object or objective.[6] "-ness" is a morphologic suffix that generates a noun from another type of word, most often from an adjective. It is used for abstract expressions of quality. The suffix is usually translated into Czech as "- (n)ost" (e.g.; *politeness* — zdvořilost), less frequently as "-ství" (e.g.; *richness* – bohatství). The literal translation of the word *aboutness* into Czech would then be "o-nost," "o čem-nost" or "o čem-ství." The Oxford English Dictionary defines *aboutness* as "the quality or fact of relating to or being about something"[7] but as we shall see in Part 3 of this article, experts may radically differ in their understanding of the concrete meaning of the term.

Czech does not have its own terminological equivalent, using instead expressions such as "topic," "content," "subject," "subject-matter," "theme," "genre," and also "meaning" and "sense." However, each one of these expressions already has its terminological use established outside of the field of information science.8 Perhaps the less used, beautifully sounding word "obsažnost" ("contentness") would be worth considering as an equivalent expression, but at present, the word connotes "richness/thickness of content." There are also the neologisms,

4 CIDOC CRM (Conceptual Reference Model) was generated by International Council of Museums – ICOM.

5 SHAW, Ryan. "Information Organization and the Philosophy of History." In: *Journal of the American Society for Information Science and Technology*. June 2013, **64**(6), s. 1100. doi:10.1002/asi.22843. ISSN 1532-2882 (Print). ISSN 1532-2890 (Online).

6 Processed pursuant to Internet Language Handbook (*Internetová jazyková příručka* [online]. Praha: Jazyková poradna ÚJČ AV ČR, ©2008–2014 [cit. 2014-03-02].Available at: http://prirucka.ujc.cas.cz/?slovo=o).

7 "*The quality or fact of relating to or being about something; Philos. (of a mental state, symbol, representation, etc.) the property of being about something (existent or non-existent).*" aboutness, n. In: *OED: Oxford English Dictionary* [online]. Oxford: Oxford University Press, © 2014 [cit. 2014-03-02].
Commercially available at: http://www.oed.com/view/Entry/247514.

8 This is also corroborated by the fact that with the exception of "content" (information) no other term is included in the Czech Terminology Database of Librarianship and Information Science (TDKIV).

"předmětnost"[9] ("subjectness") and "obsahovost" ("contentness"). Nevertheless, since no singular equivalent of *aboutness* has yet been adopted in Czech[10], we shall continue using the original English term in the rest of our article.[11]

Our study presents the results of our research into the literary meaning of "aboutness," summarizing approaches to the term in information science. The paper has three parts. The first part summarizes the issue of content analysis, presenting the concept of "aboutness." The second part presents opinions of information science experts. The third part situates the concept of "aboutness" in the FRSAD model.

2 Semiotic Triangle Concerning *Aboutness*

The title of our paper asks a question concerning the subject-matter of content analysis of documents. In short: What is content? The series of questions asked in this article's motto sketch out the reasons why experts differ in their opinions, pointing out that content-related issues are not necessarily a matter of mere semantics, but that they ensue from issues happening on the level of cognition. Hence they address the very core of the way in which man attains and represents knowledge of the world which, since Socrates' dialogue with Meno, has not yet been entirely understood. Because experts formulate their hypotheses pursuant to their affiliation to a philosophy, they use different conceptual and terminological equipment and methodologies, thus also differing in the conclusions they make. Future homogenization is not to be expected in this field because all experts would have to adopt a single philosophy, which is highly improbable.

A certain agreement exists only on the very abstract level of a so-called semantic triangle, also called reference triangle, which offers the most general perspective on the semantics of communication. The most often cited version of the triangle was published in the 1923 work of Ogden and Richards, [12]though a similar triad may be found in other works, such as with G. Frege (1848–1925)[13] a Ch. S. Peirce (1839–1914)[14]. Semiotics, linguists and cognitive scientists have traditionally used variations of the triangle[15] in order to depict the referential functioning of

9 Apparently, Russian experts have opted for this possibility since in the FRSAD model translated into Russian, *aboutness* is translated as "objecteness"(*predmetnost*) (http://www.ifla.org/files/assets/cataloguing/frsad/frsad-final-report-ru.pdf [cit. 2014-03-02]).

10 Apparently, even informal language feels that introduction of a noun expressing *aboutness* is unnecessary. So far, the popular informal Czech expressions "o čem to je" (what's it about) or, as the case may be, "o tom to je" (that's what it's about) used in the sense of "(what is) the significance/meaning of this, (what) it means, have made do with prepositional forms.

11 Similar approach may be observed in Czech texts from the discipline of Logic (e.g.; RACLAVSKÝ, Jiří. *Konstrukční vs. denotační koncepce aboutness* [*Constructive vs. Denotative Conception of Aboutness*]). 2005–2013. Draft. Available from: http://www.phil.muni.cz/~raclavsky/texty/konstrukcni_vs_denotacni_aboutness.pdf [cit. 2014-03-02])

12 OGDEN, Charles Kay a Ivor Armstrong RICHARDS. *The meaning of Meaning: A Study of the Influence of Language Upon Thought and of the Science of Symbolism*. New York: Harcourt Brace & Company, 1923. 363 s.

13 FREGE, Gottlob. Über Sinn und Bedeutung. In: *Zeitschrift für Philosophie und philosophische Kritik*. 1892, NF 100, pg. 25–50.

14 PEIRCE, Charles Sanders. *Collected Papers of Charles Sanders Peirce*. Cambridge, Harvard University Press, 1931–1958. 8 sv.

15 For details, see for example ČERMÁK, František. *Jazyk a jazykověda: přehled a slovníky (Language and the Science of Language)*. 2. dotisk 3. dopl. vyd. Praha: Karolinum, 2001, 2004, 2007. Chapter 1.52, Structure of Sign and Its Relationships (Struktura znaku a jeho vztahy), pg. 24-28. ISBN 80-246-0154-0. ISBN 987-80-246-0154-0.

language that consists in a working relationship between sign and reality — that is, sign is used to express both significance (meaning) and to denote reality.

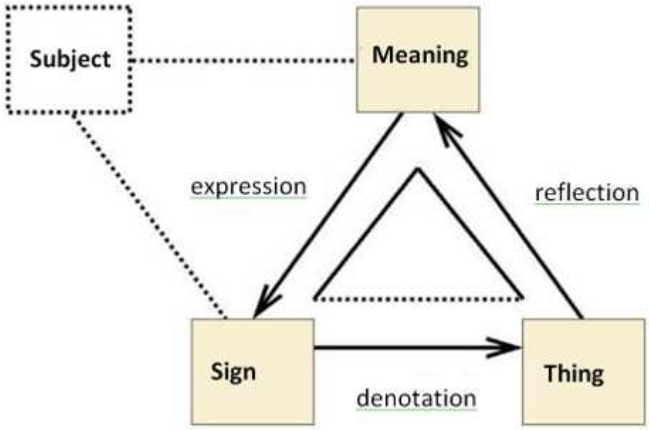

Fig. 1 Semiotic triangle

In our diagram of the semiotic triangle in Figure 1, any part of reality is represented by a working term, "thing." We could have selected another expression for "everything," such as "entity," a term popular both in information science and in the database world (e.g.; ČSN ISO TR 9007, FRBR, FRAD, FRSAD use the term), or, as the case may be,"object." However, we have chosen another term, "thing," because of its frequent use in texts of philosophy and information ontologies (e.g.; "*owl:Thing*" appears at the top of the hierarchy of OWL dictionary)[16]. Therefore, "thing" is not used here in the sense of "material object" since it will also represent abstract entities and events. The semantic relationship of "sign denoting a thing" is marked as indirect in the triangle, mediated as so-called conceptualization or representation of the meaning of reality via concepts and systems of signs.

A fourth element is at times added to the three factors at the apexes of the triangle: that of a subject or agent; that is, he who generates concepts in his mind, representing those via signs.[17] The subject adds the following into the triangle: 1. a subjective dimension; that is, a unique way in which each individual understands meaning; 2. a pragmatic dimension; that is, comparison of meaning with that which the subject intends to use.[18] The pragmatic aspect of the subject's relationship to reality as regarded by him represents the relationship of relevance between the subject's cognitive necessity and the meaning of the thing as it is becoming known to the subject; that is, the usefulness of the thing for the subject.

The triangle is most often used in linguistics that concentrates particularly on (language) sign. However, there are other domains that adopt the triangle, its concept apex being the most frequent point of interest. At present, there are three important domains where the semiotic triangle is used for the construction of models consisting of concepts: 1. conceptual models

16 Web Ontology Language

17 For example, L. Tondl used the term "observer" or, as the case may be, "language user" for subject (TONDL, Ladislav. *Problémy sémantiky (Issues of Semantics)*. Praha: Karolinum, 2006, pg. 30–31. ISBN 978-80-246-1075-7).

18 For details, see for example HUDÁKOVÁ, Miriam. Nevyhnutelnost subjektivity při pořádání informací a znalostí. ("Inevitability of Subjectivity in Organization of Information and Knowledge") In: *Ikaros* [online]. 2006, **10**(3) [cit. 2014-03-02]. URN-NBN:cz-ik3204. ISSN 1212-5075. Available from: http://www.ikaros.cz/node/3204.

designed in computer science for the development of information systems and software applications; 2. conceptual models constructed by scientists in the fields of gnoseology, psychology and cognitive sciences in order to better understand the principles of human reasoning; 3. conceptual models generated in information science as both the results and the tools for organization of knowledge.

A slightly simplified analogy presents itself between the three above-given domains of conceptual modeling and the three "words" of K.R. Popper[19] who divides objective reality into the following worlds: 1. the physical world of the matter; 2. the world of human cognition 3. the world of speech — the abstract world of knowledge and creativeness of the human spirit. There are borders between the words, but those at the same time represent points of interaction. However, Popper argues that direct interaction between World 1 and World 3 is impossible; only via World 2 can these two worlds communicate.[20] While the conceptual models of human cognition modeling World 2 are markedly individual, differing practically with each expert (e.g.; G. Lakoff[21] provides a comprehensive summary), conceptual models of World 1 and World 3 are being subjected to standardization. Model CIDOC CRM[22] that expresses *aboutness* with its P129 property as "is about (is the subject of)" represents one of the steps toward homogenization. FRSAD[23] model, simply named by its designers as "conceptual model aboutness"[24] is another example (more on this model in Part 4 of this paper). CIDOC CRM and FRSAD models are static, structural models of *aboutness.* On the other hand, ISO 5963:1985, Methods of Analysis

[19] Opinions regarding the usability of Popper's conception of his three worlds for information science differ. B. C. Brookes, for example, ranks among its defenders (BROOKES, Bertram C. "Foundations of Information Science: Part I. Philosophical aspects." In: *Journal of information science*. October 1980, **2**(3–4), 125–133. doi:10.1177/016555158000200302. ISSN 0165-5515), D. Rudd is one of its critics (RUDD, David. Do we really need World III? Information science with or without Popper. In: *Journal of information science*. August 1983, **7**(3), 99–105. doi:10.1177/016555158300700301. ISSN 0165-5515.)

[20] POPPER, Karl Raimund a John Carew ECCLES. *The Self and Its Brain*. London: Springer International, © 1977. Chapter P2, The worlds 1, 2 and 3, s. 36-50. ISBN 0-415-05898-8 (brož.). ISBN 978-0-415-05898-8 (paperback).

[21] LAKOFF, George. *Ženy, oheň a nebezpečné věci: co kategorie vypovídají o naší mysli* (*Women, Fire and Dangerous Things: What Categories Have to Say About Our Mind)*. 1st Ed. Praha: Triáda, 2006. Chaper 2. Od Wittgensteina k Roschové (From Wittgenstein to Rosch), pg. 25–66. ISBN 978-80-86138-78-7.

[22] ISO 21127:2006. *Information and Documentation – A Reference Ontology for the Interchange of Cultural Heritage Information*. 1st ed. Geneva: International Organization for Standardization, 2006. 108 s. Also freely available in electronic form at: CIDOC: http://www.cidoc-crm.org/docs /cidoc_crm_version_5.0.4.pdf [cit. 2014-03-02].

[23] In Appendix C of the Report, the authors make direct references to Ogden's a Richards's triangle. IFLA Working Group on the Functional Requirements for Subject Authority Records. *Functional Requirements for Subject Authority Data (FRSAD): A Conceptual Model*. ZENG, Marcia Lei , Maja ŽUMER a Athena SALABA, ed. Berlin: De Gruyter Saur, © 2011, pg. 48.)

[24] ŽUMER, Maja, Marcia Lei ZENG a Athena SALABA. *FRSAD: Conceptual Modeling of Aboutness*. Santa Barbara (Calif.): Libraries Unlimited, [2012]. vii, 121 s. Third Millennium Cataloging. ISBN 978-1-59884-794-9 (paperback). ISBN 1-59884-794-5 (paperback). ISBN 978-1-59884-795-6 (Online). ISBN 1-59884-795-3 (Online).

of Texts, Their Content and Selection of Indexed Terms, is the result of attempts to standardize a dynamic, processing model.[25]

However, given the high degree of abstraction of the semiotic triangle, its application onto the field of organization of knowledge has proven difficult from the start, the issue being the content of its three apexes. It has been difficult to use concrete objects of interest of information science, such as the following entities of the family of FRBR models: work, expression, manifestation, name (nomen), theme (thema). There is more than one possibility which makes the selection process difficult. The conceptual models of World 1, which are used for information systems, situate objects and events of the real world at the apex marked as "Thing," represented by concept at the apex, "Meaning" and denoted at the apex of "Sign" by a language expression or by a graphic symbol. However, it is Popper's Word 3 that is the object of knowledge organization— the abstract world of human theories and creative effort that naturally includes information resources. Without doubt, content analysis of information resources does not analyze the thing the source mentions but rather the meaning of the information registered about the thing. This is yet another reason why deciding the placement of *aboutness* in the semiotic triangle is problematic — should it go to one of the apexes (if so, which one: that of Thing, Meaning or Sign?), or one of its sides; that is, relationship (again, which one: that of Reflection, Expression or Denotation?)[26].

Perhaps, then, the concept of semiosis chains formulated by Ch. S. Peirce and called "unlimited semiosis" by Umberto Ecco[27] might provide some solution. Ch. S. Peirce used the triangle for the triadic composition of the sign, using the following expressions: "object" for thing, "representamen" for sign, and "interpretans" for meaning. "Semiosis" was then the mutual interaction of the three components of the sign. Peirce argues,

"Anything is a sign that refers something else (its interpretans) to an object to which it itself refers (its object) in the same way, whereas interpretans gradually becomes a sign, repeating this process ad infinitum."[28]

In theory, unlimited semiosis therefore makes for an endless chain of new signs to be derived from already existing signs, for example, by making statements about statements. J. Mai used this method when analyzing the process of indexing. He arrived at three interconnected semiotic triangles that gradually depict the process of document analysis in the following three steps that

[25] ČSN ISO 5963 (01 0174). *Dokumentace. Metody analýzy dokumentů, určování jejich obsahu a výběru lexikálních jednotek selekčního jazyka (Documentation: Methods of Document Analysis, Determination of their Content and Selection of Lexical Units of Selection Language)*. Praha: Český normalizační institut, 1995. 10 pg. [Czech version ISO 5963:1985]

[26] The problems mentioned are, however, only hypothetical. In actuality, no-one makes us enter key terms of knowledge organization into the Procrustes' bed of semiotic triangle. So more than anything, this is a thought experiment on our part, testing whether this semiotic model might be used as a tool for understanding the fundamentals of key terms of knowledge organization.

[27] E.g.; ECO, Umberto. *Lector in fabula: role čtenáře, aneb, Interpretační kooperace v narativních textech (Lector in Fabula: Role of Reader or Interpretation Cooperation in Narrative Texts)*. 1st Ed. Praha: Academia, 2010. Kapitola 2.8, Semióza neomezená a pragmatická, pg. 58-61. ISBN 978-80-200-1828-1.

[28] PEIRCE, Charles Sanders. Grammatica speculativa. In: PALEK, Bohumil, ed. *Sémiotika: Ch. S. Peirce, C. K. Ogden and I. A. Richards, Ch. W. Morris, H. B. Curry (Semiotics: Ch. S. Peirce, C. K. Ogden and I. A. Richards, Ch. W. Morris, H. B. Curry)*. 2nd revised edition, Praha: Karolinum, 1997, pg. 69. ISBN 80-7184-356-3.

correspond to ISO 5963:1985: 1. determination of the subject of the document (the document is the sign while meaning is the subject of the document); 2. formulation of the subject of the document via statements made in natural language (sign is the subject of the document while its language description is its meaning); 3. translation of the subject into selection language (description of the subject of the document is the sign while the subject call word or another element of the selection language is its meaning).[29]

Many experts[30] point out that Peirce's semiotic triad is too complex a model whose ambiguity is partially due to the fact that the author inconsistently defines its individual components in different parts of his work. However, we shall use the above-mentioned application of unlimited semiosis by J. Mai to produce a concept model of *aboutness* for the domain of knowledge organization where the apex of Symbol of an imaginary semiotic triangle shall be replaced with that of Information. Unlimited semiosis will make it possible to regard the results of the content analysis of the sign of *aboutness* as another sign.

The term, "information" is used here as a representative: terms such as document, file, information source, object of information, work, knowledge, knowledge artifact and also information request or, as the case may be, enquiry, may equally be projected onto the apex. What matters to us are their key components and whether these fit with our own line of inquiry; that is, information, namely its semantic dimension.

We have selected the format of class diagram in the language of UML (Unified Modeling Language)[31] for the presentation of our model which we are using in accordance with ISO/TR 24156.[32] A class symbolized by an oblong shape always represents a set of of objects with the same characteristics. While objects are perceived as instances or illustrations (exemplifications) of a class, they are not rendered in the class diagram; UML offers a special diagram of objects for those purposes. Our model uses the following types of relationships: 1. association, expressing any semantic relationship via a line ending with an arrow in cases of asymmetric association; 2. generalization which renders the hierarchy of higher entities with an triangular arrow that points at the higher class.

Figure 2 (see below) shows the general organization principle via class diagram; for better clarity, the diagram is accompanied by an object diagram that includes examples of organization of concrete things. Organization is generally defined as intentional and purposeful

[29] MAI, Jens-Erik. "Semiotics and Indexing: An Analysis of the Subject Indexing Process." In: *Journal of documentation*. September 2001, **57**(5), 591–622. doi:10.1108/EUM0000000007095. ISSN 0022-0418.

[30] For example, U. Eco argues that Peirce is "rightfully considered a contradictory thinker" (69), and elsewhere he states that "Peirce seems incoherent, momentous and contradictory" (39). Eco considers his English "totally horrific" giving a concrete example of an instance where it "grows into a kind of peculiar, defining poetry" (51). *(Lector in Fabula: Role of Reader or Interpretation Cooperation in Narrative Texts)*. 1st. Ed. Praha: Academia, 2010. Chaper 2, Peirce: Sémiotika jako základ textové kooperace (Peirce: Semiotics as Basis for Textual Cooperation), pg. 39–64. ISBN 978-80-200-1828-1.

[31] ISO/IEC 19505-1:2012. *Information Technology – Object Management Group Unified Modeling Language (OMG UML) – Part 1: Infrastructure*. 1. ed. Geneva: International Organization for Standardization, 2012. 220 s.; ISO/IEC 19505-2:2012. *Information Technology – Object Management Group Unified Modeling Language (OMG UML) – Part 2: Superstructure*. 1. ed. Geneva: International Organization for Standardization, 2012. pg. 740

[32] ISO/TR 24156:2008. *Guidelines for Using UML Notation in Terminology Work*. 1. ed. Geneva: International Organization for Standardization, 2008. pg. 19

implementation of order and structure (i.e.; defining elements and their mutual relationships) which should make a performance of a function possible.[33] Likeness or sameness is the general principle of organization of things while the usual criterion of equivalence is a set of certain characteristics of ordered objects, the so-called "isness" in English.[34] E. C. Richardson classically defines this as "connecting of similar things."[35] In Figure 2, connection-making on the basis of equivalence is rendered by the relationship of recursive association of the "Thing" class, that is, connection of the thing with itself. In object diagram, the recursive association of "things being equivalent to another thing" is instantiated or realized in the form of binary associations, linking together those instances of the "Thing" class that comply with the prerequisite of equivalence (thus we may project animals, for example, in one class and plants in the other).

Two mutually complementary processes may be detected within the general principle of organization: collocation and individualization. Collocation means that we place things that are in some way similar into one and the same set (class, category). Individualization means that at the same time, they need to be distinguishable because a single set cannot contain the same elements (that is why each instance in the picture has its own, unique name). Equivalence serves for class placement, difference helps to identify and order (sort) within the given class.[36]

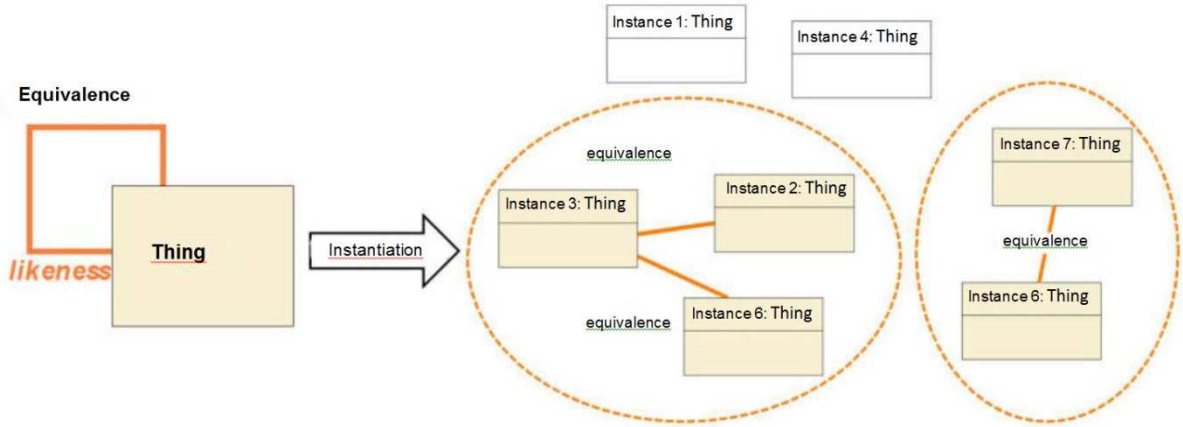

Fig. 2 General principle of organization

The diagram of general organization in Figure 3 has additional elements that are important for knowledge organization: subject and information. In view of the general definition of organization, knowledge organization may signify any intentional activity that introduces structure into existing

[33] As apparent from this definition, system approach that views system as unity of structure and function has been used for its formulation.

[34] For the purposes of this article, we are not considering alternative criteria or equivalence for categorization, e.g.; theory of prototypes, family resemblance, etc. Detailed listing is given in LAKOFF, George. *Ženy, oheň a nebezpečné věci: co kategorie vypovídají o naší mysli* (*Women, Fire and Dangerous Things: What Categories Have to Say About Our Mind)*. 1st Ed. Praha: Triáda, 2006. ISBN 978-80-86138-78-7.

[35] *"Classification is in its simplest statements putting together like things."* RICHARDSON, Ernest Cushing. *Classification: Theoretical and Practical.* Together with an appendix containing an essay towards a bibliographical history of systems of classification. New York: Charles Scribner's Sons, 1901, pg. 1. ark:/13960/t6m041g07.

[36] We find in thus formulated principle of equivalence a possible answer to one of Meno's questions from the introduction to this article: information search does not concern achievement of identity but of resemblance of search question/request and response/information.

knowledge in order to mediate better access to it. In other words, the subject recognizes and identifies important elements of existing knowledge, defining their mutual relationship, while access to knowledge is the function that makes the selected order possible. These activities underpin nearly the entire life circle of knowledge, from representation (expression), potential communication and sharing, registering, publishing, processing (identification, description, content analysis, indexing, transformation) to depositing. However, knowledge in itself is not the unit of knowledge organization, being a part of World 2; it is rather its record; that is, the source of information belonging to World 3, represented by "Information" in our diagram.[37]

As mentioned before, the working term, "Thing" in our model represents any part of reality, which includes information. The symbol of generalization/genesis in Figure 3 highlights the fact that information is a case-specific type of thing that has the general characteristics of things, while differing at the same time from all other things by its specific properties. Subject (yet again, a specific type of thing) that holds any kind of relationship to information, is another part of reality that interests us here. A subject may be the author of the information, its processor (indexer) and also its user. For the sake of completeness, a class of "Other Things," which are considered neither information nor subject, is also included in the diagram.

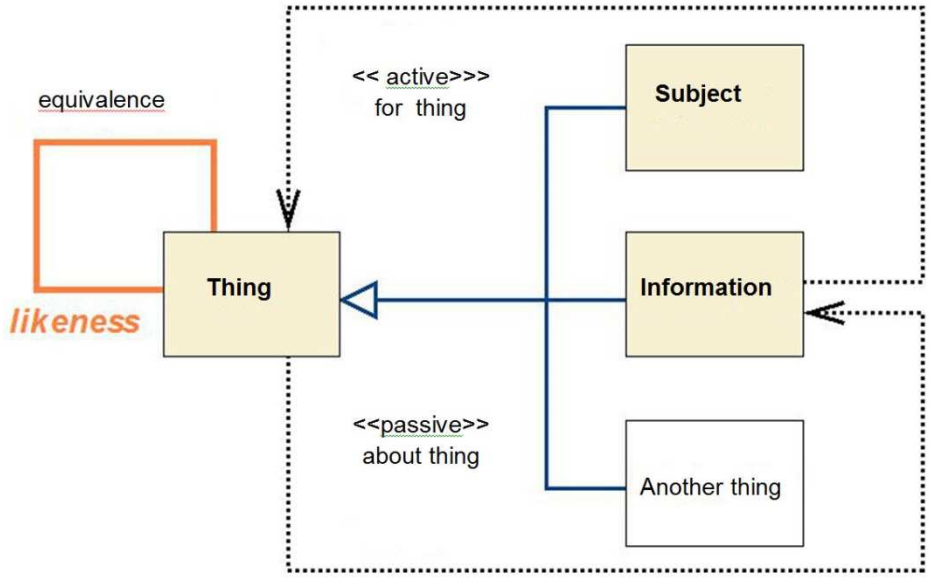

Fig. 3 Information as a specific kind of thing

Naturally, the principle of equivalence for organization of things is equally applicable for organization of information as specific instances of things.[38] However, this is where we run into problems since information, from the point of organization, differs from all other things by the fact

[37] For more detailed account of the framework of knowledge organization, see the following article: KUČEROVÁ, Helena. České termíny pro věcné pořádání informací po 40 letech: příspěvek k terminologické diskusi ("Czech Terms for Factual Indexing of Information After 40 years: Contributing to Discussion on Terminology). In: *ProInflow: časopis pro informační vědy* [online]. 2013, **5**(Speciál), [cit. 2014-03-02], pg. 1–19. ISSN 1804-2406. Available at: http://pro.inflow.cz/ceske-terminy-pro-vecne-poradani-informaci-po-40-letech-prispevek-k-terminologicke-diskusi nebo http://pro.inflow.cz/sites/default/files/pdfclanky/ProInfow_Kucerova_final_0.pdf.

[38] In generalization relationship, there is inference via genetic transmission of characteristics and relationship of superior unit.

that the criterion of equivalence, which is based on constitutive inner properties, is barely used for information organization. Rather than asking "what is" information (which is the ontological level, conceived as *isness* in English, e.g.; a document "is," a manuscript "is," the International Decimal Classification "is"), we ask "what is it about" (which indicates a gnoseological level conceived as *aboutness*, e.g.; a document "is about" manuscripts, it "is about" the International Decimal Classification).

In our model, information is purely purpose-oriented because of its role in knowledge organization. However, E. Svenonius argues that "the way in which information is defined determines the what and the how of its organization…, for example, a work is difficult to define because this means that information must also be defined." Svenonius then identifies the consequences of the difficulties regarding definition of information by stating that "what is difficult to identify is equally difficult to describe, and so it is also difficult to organize."[39] As regards information, this makes us narrow down our focus somewhat to the following areas: 1. ontological — information about the thing; 2. pragmatic — information for the thing. The areas are schematically rendered in Figure 3 as <<passive>> (information that reflects reality) and <<active>> (information that changes reality). Our conception is broader than that of Shannon's communication model of information that intentionally excludes meaning of the message. In our model, on the other hand, a gnoseological aspect is intrinsic to both the ontological and the pragmatic conception of information; that is, we conceive of information as an entity that has discernible meaning (sense).

1. From the point of view of ontology, information is perceived as message. "Reflection" is the key term here through which information may be defined as something that is a reflection (i.e.; reflection, model, representation) of reality.[40] This point of view is passive and static, conceiving information as the end-result of a process. The starting premise here is that information is always "about something." In order for information to exist, then, there has to exist "something;" information then reflects this "something" and reports on it. However, reflection does not reproduce reality in its totality but rather selects the unusual and the dissimilar by which it differs from its surroundings. Diversity and variety, then, may not necessarily concern only the external appearance of a reflection in a mirror— not merely reflecting the "surface" — but mainly the internal structure, organization and configuration of the given phenomenon.[41]

In Figure 3, the "information is about a thing" relationship is an indirect recursive association that makes it possible for information to become the subject of representation itself; this may be useful, for example, when considering relationships between information resources and metadata.

2. The pragmatic point of view understands information as an injection of a shape; as shape-formation ("in-formation"). The pragmatic approach to information is underpinned by the terms of "entropy" and "relevance." This point of view is active and dynamic: information is perceived as activity and process of affecting reality that results in increased organization and order, and,

[39] SVENONIUS, Elaine. *The Intellectual Foundation of Information Organization*. Cambridge (Mass): MIT Press, 2000, pg. 12–13. ISBN 0-262-19433-3. ISBN 978-0-262-19433-4.

[40] "If there are changes happening in the object that reflect the working of another object, then the first object becomes information carrier about the second object. Information. In: *Filozofický slovník*. Praha: Svoboda, 1976, pg. 195.

[41] This view points out that organizational principle is already present in the mere fact of existence of information.

consequently, lower degree of entropy. However, while the relationship of information to reflected reality is the point of interest in the category of reflection, we are here examining the relationship of information to its user/receiver/end point (information is "for someone/something"). We use the term, "relevance" for this relationship; information is relevant if it is complete, adequate and new to its user, thus complying with his (information) need. Naturally, the pragmatic scenario actively involves the subject, as seen in Figure 4 (see below).

In theory, we may situate *aboutness* in the following three places in our model. 1. *Aboutness* may be viewed as a symmetrical relationship to equivalence, that is to say, as yet another kind of recursive association of instances of things (in the sense of "things are about things"). However, this would mean that everything has aboutness while in fact, only World 3 objects may be "about something." Those are represented by the class of "Information" in our diagram. *Aboutness* thus may be the characteristic of information, the ramifications of which are twofold: 2. *Aboutness* as the relationship of information—thing (depicted in Figure 3 as the associative relationship of "information is about things"). However, *aboutness* is different from reflection and description: while every thing may be described, *aboutness* may only be attributed to some things. If this relationship were valid, aboutness might equal a static conception of information, or, as the case may be, aboutness might equal the reflection of reality, which is obviously untrue. Therefore, the third possibility seems the most viable: 3. *aboutness* as cognitive construction generated by the subject, situating it in relationship with information that "is about it" or, as the case may be, "has it" or relates to it (expert opinions on this matter will be presented in Part 3 of this paper).

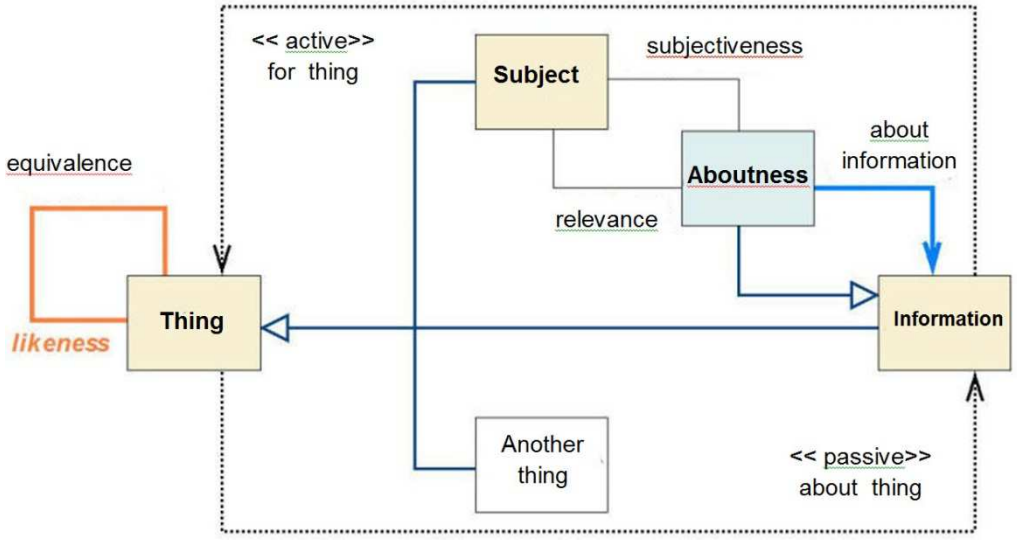

Fig. 4 Relationship among information, subject and *aboutness*

The symbol of generalization in Figure 4 graph therefore renders *aboutness* as a specific kind of information while at the same time, the indirect recursive association, "about information," signals that a message about this information is being transmitted. The association with subject suggest the dual effect subject has on the definition of *aboutness*: subjectivity involved in content understanding is underpinned by the individual nature of the cognitive processes of each individual, while relevance points out that rather than assessing the source content objectively, the subject relates it to the usefulness of the content and to his own or another subject's information needs. The relationships in Figure 4 further imply that *aboutness* does not encompass all aspects of the relationship of information to reality phenomena, nor does it encapsulate all the possibilities of their organization.

3 *Aboutness* In Literature of Information Science

Aboutness appeared in information science literature at the end of the nineteen-sixties for the first time (1968 P. Wilson,[42] 1969 R. A. Fairthorne[43]). But it was the years 1977 and 1979 that signified the turning point, because a symposium on this topic was organized by the British ASLIB Co-ordinate Indexing Group CIG in 1977, and a conference [44]was organized by ASLIB and BCS (British Computer Society) in 1979. In our lands, B. Kovář tackled this issue as early as in 1974. His perspective is multifaceted, and even though he does not support his claims with evidence from research, he essentially covers all significant aspects of the term. Kovář is interested in *aboutness* as the result of content analysis, using expressions such as "content," "subject," or, as the case may be, "theme," and defining *aboutness* as a set of words of natural language that formulate the content of document, its parts, or individual information the document contains. This set of words is used for the formulation of the texts of abstracts and for subject entries, descriptors, and classification signs. Kovář argues that besides content itself, perspective and the form in which theme is delivered is equally crucial. He points out the meaning of context outside of the document itself, which may be other documents, authors and other field experts. Therefore, Kovář cautions that anticipated users must be taken into account, and that the granularity (depth) of content analysis must be adjusted accordingly.[45]

We have mentioned that *aboutness* is far from being uniformly understood in information science. B. Hjørland even argues that *aboutness* brings nothing new, being just another vague term used by some experts to replace the poorly defined term "subject."[46] In fact, Hjørland attempted to define the term "subject" himself, having prepared a typology of "subject" (content) which contained the following five notions: 1. naive notion of the subject — identification of subject with its name, such as document content being derived from the words of its title; 2. subjective idealism — content is what the author/reader/librarian believe the content of the document is (the content is "in the head"); 3. objective idealism — content objectively exists, and may be determined via professional and scientific methods (content is "in the document"); 4. pragmatic notion of content — content is perceived as a tool with which a goal may be achieved, and so its definition is produced with the goal in mind; 5. realistic/materialistic notion of content — definition of the key characteristics of document precedes its content analysis. The author favors the last notion, proposing to use epistemological (gnoseological) methods for its application.

[42] WILSON, Patrick. *Two Kinds of Power: An Essay on Bibliographical Control*. Berkeley: University of California Press, 1978, © 1968. pg. 155.

[43] FAIRTHORNE, Robert A. "Content Analysis, Specification, and Control." In: *Annual review of information science and technology.* Vol. 4, 1969. Carlos A. Cuadra, ed. Chicago: Encyclopaedia Britannica, 1969, pg. 73–109.

[44] MacCAFFERTY, Maxine a K. GRAY, ed. *The Analysis of Meaning: Informatics 5:* proceedings of a conference held by the Aslib Informatics Group and the BCS Information Retrieval Specialist Group, 26–28 March 1979, the Queen's College, Oxford. Aslib, 1979. 302 s. **ISBN** 978-0-85142-125-4. **ISBN** 0-85142-125-3.

[45] KOVÁŘ, Blahoslav. *Obsahová analýza dokumentu (Content Analysis of Documents)*. Praha: ÚVTEI, 1974, pg. 18–21. Metodický leták, sv. 105.

[46] HJØRLAND, Birger. "Towards a Theory of Aboutness, Subject, Topicality, Theme, Domain, Field, Content … and Relevance." In: *Journal of the American Society for Information Science and Technology*. July 2001, **52**(9), 774–778. doi:10.1002/asi.1131. ISSN 1532-2882 (Print). ISSN 1532-2890 (Online).

Hjørland further argues that determining the potential of the subject for knowledge enhancement as exactly as possible is the ultimate goal of subject analysis of documents. [47]

J. Furner[48] distinguishes between *aboutness* and the so-called *subjecthood* (defined by the Oxford English Dictionary as "the state or condition of being a subject"[49]), pointing out that the distinction is not purely theoretical because it impacts the designing of library catalogues, for example. *Aboutness* is the response to questions about the logical nature of the relationship between statement/document and the thing we claim to be the statement/document. *Subjecthood* defines the ontological nature of the thing we claim to be the statement/document. Therefore, *aboutness* is the relationship to the subject of the work, while *subjecthood* is the subject itself. Like B. Hjørland, Furner identifies various notions of *aboutness*, of which the following two, termed by Furner as "realism" and "nominalism," stand out. Realism considers *aboutness* to be the attribute of a work, while nominalism argues that aboutness is the relationship between the work and theme. The author maintains that bibliographic classification systems and document indexing are based in no other but realistic approach, regardless of the scarcity of attention that specialized literature pays to this approach. In order to promote the nominalist model, the author proposes that subjects ought to be viewed analogically as works in FRBR model, that is, as products of creative activity. "Inasmuch as work is realized via expression embodied by manifestation, and illustrated by items, we may reasonably assume that subject (content) is realized via notions embodied by terms (term-type) and illustrated by the concrete existence of terms (term-token)?"[50] The question mark suggests that the author intends to prompt discussion rather than making a definite conclusion.

C. Beghtol produced another important study on *aboutness* in her article that examines the relationship between theories of bibliographic classification and textual linguistics, whose method of intertextuality was used for the analysis of *aboutness* and the cognitive processes of document classification.[51] The author refers to the textual linguist, T. A. van Dijk, and his theories of textual discourse that distinguish between *aboutness* and *meaning*. Beghtol maintains that while the meaning of a message of a document may differ with each subject, *aboutness* is fixed — each document has its relatively permanent objective content that does not change in time. She goes on to argue that "recognition of the relatively permanent quality of *aboutness* in documents is one of the traditional premises of bibliographic classification systems."[52] The author applies intertextuality as a relationship of a text to another text to the relationship of text—metatext as

[47] HJØRLAND, Birger. "The Concept of 'Subject' in Information Science." In: *Journal of documentation*. September 1992, **48**(2), 172–200. doi:10.1108/eb026895. ISSN 0022-0418.

[48] FURNER, Jonathan. "FRSAD and the Ontology of Subjects of Works." In: *Cataloging & classification quarterly*. 2012, **50**(5–7), 494–516. ISSN 0163-9374. doi:10.1080/01639374.2012.681252.

[49] "The state or condition of being a subject." subjecthood, n. In: *OED: Oxford English Dictionary* [online]. Oxford: Oxford University Press, © 2014 [cit. 2014-02-02]. Commercially available at: http://www.oed.com/view/Entry/192696.

[50] FURNER, Jonathan. FRSAD and the Ontology of Subjects of Works. In: *Cataloging & classification quarterly*. 2012, **50**(5–7), s. 513. ISSN 0163-9374. doi:10.1080/01639374.2012.681252.

[51] BEGHTOL, Clare. "Bibliographic Classification Theory and Text Linguistics: Aboutness Analysis, Intertextuality and the Cognitive Act of Classifying Documents." In: *Journal of documentation*. 1986, **42**(2), 84–113. ISSN 0022-0418.

[52] BEGHTOL, Clare. "Bibliographic Classification Theory and Text Linguistics: Aboutness Analysis, Intertextuality and the Cognitive Act of Classifying Documents." In: *Journal of documentation*. 1986, **42**(2), s. 85. ISSN 0022-0418.

follows: primary text (original) — secondary text (abstract) —tertiary text (expression of selection language). Attention is paid particularly to intertextual relationships on the level of primary texts (documents) in the same category of the same classification system, and on primary level (document) — classification (also a primary text).

W. J. Hutchins also utilizes textual linguistics. He proposes an approach to indexing based in linguistic analysis of textual structure, offering simultaneously an alternative to traditional summarizing (semantic condensation) of document content in the form of an call name in the index, for example, or in the form of key words that, however, do not always work. From the conceptual apparatus of textual linguistics, he selects terms such as "theme" (the starting point, the given, that which the author assumes to be known to the reader) and "rheme" (that which is new, expected by the reader), extrapolating those from the level of syntax to the level of entire texts. Two kinds of user needs are defined as follows: 1. to learn (much) about the domain of which the user knows little, which necessitates a theme rather a summary of content as a whole;[53] 2. to learn something new about the domain that is very familiar to the user, in which case the method of summarizing is apt. The author concludes that more often than not, it is "theme" that is better used in the capacity of *aboutness.*[54]

In his detailed survey of content analysis, R. A. Fairthorne compares different disciplinary approaches to *aboutness.* He maintains that while logic and linguistics examine *aboutness* as an attribute of statement, being most frequently viewed as a part of its meaning, information science examines *aboutness* as the relationship between statement on the one hand, and the information needs of its recipient on the other, hence as an agent/tool with which to meet information needs. The author outlines two kinds of *aboutness*: 1. intentional *aboutness,* determining what the entire document is about; 2. extensional *aboutness* that deals with what individual parts of the document are about, simultaneously pointing out that a totality is not a mere sum total of its own parts.[55]

At about the same time, M. E. Maron comes up with a behavioral approach to *aboutness*, proposing the following kinds of "*about*": *S-about* is a subjective (non-behavioral), psychologically individual understanding of the content of a document. *O-about* is objective and behavioral, representing that which the individual user shall requests (the term he will enter into the search form when doing his search) when looking for information the document contains. *R-about* is the objective and behavioral probability of that which all potential users (e.g.; library users) shall request, meeting their information needs.[56]

B. Boyce similarly links *aboutness* to relevance, defining two kinds of *aboutness*. Theme-based *aboutness* is a necessary but insufficient condition of relevance. The so-called informativeness, that is, understandability (accessibility) and novelty to the user represents the second kind. While theme-based *aboutness* exclusively affects the completeness of the search

53 By the way, here we find yet another possible answer to Meno's questions — we ask about the unknown by using the known to ask.

54 HUTCHINS, W. John. "The Concept of 'Aboutness' in Subject Indexing." In: *Aslib proceedings*. January 1978, **30**(5), 172–181. doi:10.1108/eb050629. ISSN 0001-253X.

55 FAIRTHORNE, Robert A. Content Analysis, Specification, and Control. In: *Annual review of information science and technology.* Vol. 4, 1969. Carlos A. Cuadra, ed. Chicago: Encyclopaedia Britannica, 1969, pg. 73–109.

56 MARON, M. E. "On Indexing, Retrieval and the Meaning of About." In: *Journal of the American Society for Information Science*. January 1977, **28**(1), 38–43. doi:10.1002/asi.4630280107. ISSN 0002-8231 (Print). ISSN 1097-4571 (Online).

results, informativeness affects both completeness and exactness. Its introduction in the search process might therefore equalize the reciprocal proportion of completeness and exactness.[57]

P. Bruza and collective present an approach to *aboutness* underpinned by logic. Using formal logic, the collective has put together a common sense set of characteristics of *aboutness* and of its opposite, *nonaboutness*, in order to make *aboutness* useable both in models and algorithms of machine information search and retrieval, and from the perspective of human cogitation. The characteristics are as follows: reflexivity, asymmetry, consistence, semantic restriction, monotony, completeness, reliability.[58]

Experts in the field of information science usually concentrate on non-fictional, specialized texts rather than on fiction. A. M. Pejtersen is an exception to the rule. Having analyzed approximately 300 recorded conversations between librarians and authentic users of Danish public libraries in 1973 and 1976, the author proposes that *aboutness* be understood multi-dimensionally, pursuant to the relationship between the characteristics of documents and user needs. Four dominant dimensions of the users' notion of what a book is about have been identified: 1. content, theme, (plot, sequence of events, development of characters, description, social relationships); 2. time frame (past, present, future), place (geographic, societal, occupation); 3. author's intention (emotions, lesson, information); 4. accessibility (readability, physical characteristics). On the basis of thus identified dimensions, the author designed a multidimensional diagram of classification of fiction.[59]

Swift et al. propose a similar multidimensional principle for indexing systems which allows for different ways in which to generate document characteristics in the field of social sciences.[60]

An article by R. Shaw[61] represents one of the latest contributions to the debate on *aboutness*. Shaw applies various approaches from the history of philosophy where the point of view is often more important than the theme, distinguishing reference from presentation. He points out that historical texts (and, apparently, other texts as well), represent rather than make references to something. Therefore, he argues that models of discourse about history are more important for historians than the traditional library classification of historical places and periods, considered as the a-priori given "landscape of objects." His approach corresponds to that of H. Albrechtsen who, twenty years before Shaw's theory, proposed to adopt domain-specific approach to indexing, that is, adapting the methodologies and techniques of indexing to the particular domain in which indexing is being done.[62]

---

57 BOYCE, Bert R. "Beyond Topicality: a Two Stage View of Relevance and the Retrieval Process." In: *Information processing and management*. 1982, **18**(3), 105–109. doi:10.1016/0306-4573(82)90033-4.

58 BRUZA, P. D., D. W. SONG a K. F. WONG. "Aboutness from a Commonsense Perspective." In: *Journal of the American Society for Information Science*. 2000, **51**(12), 1090–1105. doi:10.1002/1097-4571(2000)9999:9999<::AID-ASI1026>3.0.CO;2-Y. ISSN 0002-8231.

59 PEJTERSEN, Annelise Mark. "The Meaning of 'About' in Fiction Indexing and Retrieval." In: *Aslib proceedings*. May 1979, **31**(5), 251–257. ISSN 0001-253X.

60 SWIFT, D. F., V. WINN a D. BRAMER .'Aboutness' as a Strategy for Retrieval in the Social Sciences." In: *Aslib proceedings*. 1978, **30**(5), 182–187. ISSN 0001-253X.

61 SHAW, Ryan. "Information Organization and the Philosophy of History." In: *Journal of the American Society for Information Science and Technology*. June 2013, **64**(6), 1092–1103. doi:10.1002/asi.22843. ISSN 1532–2882 (Print). ISSN 1532-2890 (Online).

62 ALBRECHTSEN, Hanne. "Subject Analysis and Indexing: From Automated Indexing to Domain Analysis." In: *Indexer*. 1993, **18**(4), 219–224. ISSN 0019-4131.

S. Shatford Layne concretizes the domain-specific approach to *aboutness,* proposing a typology of attributes for the indexing of images. The author divides the attributes into the following four categories: 1. "biographic" (attributes concerning the creation/origin of the image, attributes describing its life/changes); 2. subject-based; 3. illustrative (e.g.; the image is a poster, not an image of a poster); 4. relationship-based (image — image, image — text, image — object). Naturally, the category of subject proves to be the most problematic, and so the author further considers the following aspects: 1. the image may simultaneously be "*of*" (this time, not in the sense of "about" but "of what" — what is in the image, what is depicted; for example, a crying woman) and "*about,"* that is, what the image is about (e.g.; sadness). 2. The image in the sense of "*of*" is general and specific at the same time (e.g.; there is a crying woman in the image/in the image, there is a crying Dora Maar). 3. Four facets of content are suggested: time, space, activities and events, and objects. It is possible to combine any facet of the content with any of the above-given aspects, that is, with the specific "*of*," the general "*of*," and with "*about*."[63]

4 *Aboutness* in FRSAD Model

Being the latest of the family of FRBR models, conceptual model FRSAD (*Functional Requirements for Subject Authority Data)* was presented in 2010. The text of the report, accessible at the webpage of IFLA Committee on Standards, was published by De Gruyter Saur in 2011.[64] The report was revised in 2012 and expanded by additional examples.[65] In his 2012 article, J. Furner, member of FRSAD Advisory Group, provides a detailed account of how the Working Group approached the building of the model (the article is mentioned in Part 3 of this paper).[66]

FRSAD was generated upon requests of the community of FRBR users that asked for further analysis of entities of the third group that serves as subjects of works: concept, model, object, event, place.[67] Surprisingly enough, the analysis resulted in reduction of the number of entities, and in narrowing down the of the intentionality of the concept of *aboutness* used in the model. Members of the Working Group made their decision after a pilot study was conducted, whose

63 LAYNE, Sara Shatford. "Some Issues in the Indexing of Images." In: *Journal of the American Society for Information Science*. 1994, **45**(8), 583–588. ISSN 0002-8231.

64 IFLA Working Group on the Functional Requirements for Subject Authority Records. *Functional Requirements for Subject Authority Sata (FRSAD): a Conceptual Model*. ZENG, Marcia Lei, Maja ŽUMER a Athena SALABA, ed. Berlin: De Gruyter Saur, © 2011. vi, 74 s. IFLA series on bibliographic control, vol. 43, ISSN 1868-8438. ISBN 978-3-11-025323-8 (Print). ISBN 3-11-025323-2 (Print). ISBN 978-3-11-026378-7 (Online). ISBN 3-11-026378-5 (Online). Dostupné také z: http://www.ifla.org/node/5849 [cit. 2014-03-02].

65 ŽUMER, Maja, Marcia Lei ZENG a Athena SALABA. *FRSAD: Conceptual Modeling of Aboutness*. Santa Barbara (Calif.): Libraries Unlimited, [2012]. vii, 121 s. Third Millennium Cataloging. ISBN 978-1-59884-794-9 (paperback). ISBN 1-59884-794-5 (paperback). ISBN 978-1-59884-795-6 (Online). ISBN 1-59884-795-3 (Online).

66 FURNER, Jonathan. FRSAD and the ontology of subjects of works. In: *Cataloging & classification quarterly*. 2012, **50**(5–7), 494–516. ISSN 0163-9374. doi:10.1080/01639374.2012.681252.

67 The FRBR study voices the very proposal: "Further analysis has to be made particularly in the domain of entities of factual authorities, i.e.; thesauri and classification schemata and on relationships among those entities." *Funkční požadavky na bibliografické záznamy: závěrečná zpráva* (*Functional Requirements Concerning Bibliographic Records: Final Report*) [online]. 1st. Ed. Praha: Národní knihovna ČR, 2002, pg. 4. ISBN 80-7050-400-5. Available at: http://www.webarchiv.cz/files/dokumenty/reference/frbr.pdf [cit. 2014-03-02]

participants were asked to sort 3 000 terms into six categories derived from the subject units of FRBR. The results of the study showed that there was no agreement; the participants had a hard time distinguishing general terms from concrete terms, while a large number of the terms were placed into the category of "other."[68] For this reason, the creators of FRSDA recommended that the following entities: "concept," "model," "object," "event," and "place" were removed from the FRBF family of models, to be replaced with the generally defined entity of "thema" that may be analyzed and further divided into sub-entities by each user according to his needs (being at the same time able to continue using the original, FRBF entities if they are adequate for his needs).

The authors claim that they do not hold any philosophical stance regarding *aboutness*, seeing it rather from the user's point (however, J.Furner points out that this attitude may already constitute a philosophical stance).[69] The model comprises three entities: "work," adopted from the FRBR model, and the newly defined entities of "thema" and "nomen." The starting premise for the model argues that *aboutness* relates to work as an FRBR entity, that is to say, an abstract entity that represents a certain intellectual or artistic creation which is realized through expression, embodied through manifestation, and exemplified by item, the relationship being that of work and its subject, called "has a subject" (already in FRBR), and defined as "many — to — many" relationship.

Therefore, a single super-class called *thema* has been allocated to subject of work in the model, representing anything that may become subject of work. The authors further delineate *aboutness* and *ofness* (used, for example, for descriptions of what is depicted in pictures) as opposed to *isness* (form, genre, and target audience of resource and other characteristics that are considered attributes of work and defined as such in FRBR). FRSAD therefore conceives of *aboutness* as no more than a summary of content/theme. Compared with the theories summarized in Part 3 of this paper, this might seem a one-sided and limited point of view that does not take full account of the semantic, and — most important —the pragmatic aspects of *aboutness*.

Name ("nomen") is the third entity, which may be any sign or sign sequence naming or denoting thema. The relationship between thema and nomen is that of "has name" which is again defined as "many —to — many" relationship.

Ease of use is the main advantage of FRSAD, being general enough to fit almost any content. Instead of sets of possible, explicitly named subjects, the models makes it possible for anyone to define any range of themes via a "type of thema" attribute value of the thema entity. This simple a structure has another advantage since it is unimposing, and yet it provides a stable platform for other development. As demonstrated by concrete examples in Appendix C of the report, generalization of FSAD model makes for compatibility and mapping with other conceptual models — syntactic schemata of SKOS and OWL script, and the abstract DCMI model in particular. With more concrete examples, Appendix D then documents the applicability of FRSAD to all systems of knowledge organization.

[68] IFLA Working Group on the Functional Requirements for Subject Authority Records. *Functional requirements for subject authority data (FRSAD): a conceptual model*. ZENG, Marcia Lei, Maja ŽUMER a Athena SALABA, ed. Berlin: De Gruyter Saur, © 2011, s. 41.

[69] FURNER, Jonathan. "FRSAD and the Ontology of Subjects of Works." In: *Cataloging & classification quarterly*. 2012, **50**(5–7), s. 496. ISSN 0163-9374. doi:10.1080/01639374.2012.681252.

However, the general character of FRSAD is also its disadvantage, as inconsistent partial solutions may be produced during the specification process of a particular use.

5 Conclusion

As syntactic issues concerning electronic communication have been resolved and attention is now paid to the higher level of semantics, we have the opportunity to revisit the half-forgotten, pre-internet visions that were before, from the nineteen fifties to the nineteen seventies, viewed as purely theoretical; these may now be made reality with our new technologies.

Our summary of different expert opinions on the issue of *aboutness* in Part 3 of this article has demonstrate that many assume that semantic activities performed on Word 3 objects are realized with Word 2 tools, that is to say, that the content of information resources is accessible via human cognition. Until recently, we assumed that *aboutness* must simply be thought up. The interesting question is now how to proceed when assigning work with content to machines, such as software agents in semantic web.[70] One way to do this is to try and match machine semantic processes with the principles on which the human mind operates, as suggested by Vannevar Bush in his 1945 design of Memex,[71] a personal information system, and as is being done these days by technologies based in neural networks, for example. Another alternative might be the idea to replace human cognitive processes with technologies that use their own effective methods, such as statistical and mathematic modeling,[72] instead of copying the human processes. These and similar issues are now the domain of artificial intelligence studies.

It is beyond doubt that we benefit from contemporary information technologies that make it possible to "materialize" some of the so-far abstract phenomena. Only in the environment of relation databases and SQL language can we literally "touch" the sets and results of Boolean logical operations. Even on the level of semantics, ontologies and their languages, RDF and OWL in particular, now yield similar effect. To be sure, preparation of data for "materialization" of content is far more complex than that on the level of syntax in database systems. In 1969, R. A. Fairthorne made the following comment about the quintessential limitations of information system that still holds today:

"No information system can process pure knowledge. The only available knowledge within its competence is knowledge of the (recorded) discourse, not knowledge of what the given discourse is about. …No information system can replace the individual user when selecting

70 Among others, a survey of methodologies and technologies of indexing is provided by J. Schwarz (SCHWARZ, Josef. Současný stav a trendy automatické indexace dokumentů: přehledová studie. ("Current State and Trends of Automatic Indexing of Documents: Survey Study.") In: *Ikaros* [online]. 2003, **7**(3) [cit. 2014-03-02]. urn:nbn:cz:ik-001300. ISSN 1212-5075. Available at: http://www.ikaros.cz/node/1300.)

71 V. Bush was referring to the psychological theory of associationism that explains physical phenomena through their physical manifestation in the form of basic elements of consciousness (e.g.; senses and imagination) that are formed into larger entities via association.

72 "When man wanted to imitate walking, he invented the wheel, which does not at all look like a leg." APOLLINAIRE, Guillaume. *Prsy Tirésiovy: nadrealistické drama o dvou jednáních s prologem (The Breast of Tiresias):* Praha: Odeon, 1926, pg. 10.

documents. … Librarians and library system are no replacement for the reader but merely his agents and tools."[73]

Everything seems to suggest that similarly to communication across time, communication across different science disciplines will prove very useful. For example, logic examines the *aboutness* of individual clauses/utterances which, until recently, was not applicable in information science and in knowledge organization in particular because these focus on documents — on entire texts. Nowadays, with the onset of semantic web and interconnected data, knowledge of different kinds of logic becomes the basis for construction of languages that process interconnected data. Moreover, logic usually uses *aboutness* to determine the veracity of a statement, which corresponds with the proposed layer of "trust" of semantic web.[74]

"Anyone can write anything about anything and any future reader may use it for anything else."[75] In his statement, R. A. Fairthorne succinctly contends that *aboutness* depends equally on the content of the message and on how the content will be used. His prophetic conception of an open world is now fully explored in the environment of semantic web.

[73] FAIRTHORNE, Robert A. Content Analysis, Specification, and Control. In: *Annual review of information science and technology.* Vol. 4, 1969. Carlos A. Cuadra, ed. Chicago: Encyclopaedia Britannica, 1969, pg. 74.

[74] One of the numerous diagrams rendering the layers of semantic web, available at: http://www.w3.org/2007/03/layerCake.png [cit. 2014-03-02].

[75] FAIRTHORNE, Robert A. "Content Analysis, Specification, and Control." In: *Annual review of information science and technology.* Vol. 4, 1969. Carlos A. Cuadra, ed. Chicago: Encyclopaedia Britannica, 1969, pg. 84.