Checking Knowledge in Online Encyclopaedias: Towards a Behavioural Approach to Data Accessibility

Marcin Overgaard Ptaszynski, *Centre for Lexicography, Aarhus School of Business, Aarhus University, Denmark (maop@asb.dk)*

Abstract: This article outlines a new approach to data access in electronic encyclopaedias. Contrary to most of the existing research, which treats access as a lexicographer-dominated, structural concept, access is viewed here as a process which is essentially user-driven. The role of the lexicographer is to facilitate this process, so that it can become rapid and unimpeded, but their control of it should be reduced to a minimum. This can only be achieved if access to data is adjusted to the specific type of need for information. Accordingly, the article examines three forms of data access applicable in situations in which the user experiences the need for a finite, specific amount of information. The results of the examination confirm that access to data is indeed a process whose success depends on the interaction between the user, the data, and the type of need for information which the given reference work aims to satisfy.

Keywords: ENCYCLOPAEDIA, ELECTRONIC LEXICOGRAPHY, ACCESS, STRUCTURE, ACCESS ROUTE, SIMPLE INFORMATION NEED, COMPLEX INFORMATION NEED, KNOWLEDGE ACQUISITION

Opsomming: Die nagaan van kennis in aanlynensiklopedieë: Op weg na 'n gedragsbenadering tot datatoeganklikheid. Hierdie artikel skets 'n nuwe benadering tot datatoegang in elektroniese ensiklopedieë. In teenstelling met die meeste bestaande navorsing wat toegang behandel as 'n leksikograaf beheerste, strukturele konsep, word toegang hier gesien as 'n proses wat hoofsaaklik gebruikersgedrewe is. Die rol van die leksikograaf is om hierdie proses te vergemaklik sodat dit vinnig en ongehinderd kan plaasvind, maar sy beheer daarvan behoort tot 'n minimum beperk te word. Dit kan slegs bereik word indien toegang tot data aangepas word by die spesifieke soort behoefte aan inligting. Gevolglik behandel die artikel drie maniere van datatoegang wat van toepassing is in situasies waarin die gebruiker die behoefte vir 'n definitiewe, spesifieke hoeveelheid inligting ondervind. Die resultate van die ondersoek bevestig dat toegang tot data inderdaad 'n proses is waarvan die sukses berus op die wisselwerking tussen die gebruiker, die data en die soort behoefte aan inligting wat die bepaalde naslaanwerk beoog om te bevredig.

Sleutelwoorde: ENSIKLOPEDIE, ELEKTRONIESE LEKSIKOGRAFIE, TOEGANG, STRUKTUUR, TOEGANGSROETE, EENVOUDIGE INLIGTINGSBEHOEFTE, KOMPLEKSE INLIGTINGSBEHOEFTE, KENNISVERKRYGING

1.   Introduction

What is the wingspan of an albatross? Anyone who would like to know the right answer may ask an expert in the field, e.g. an ornithologist or a biologist. If none of them happens to be available and the issue needs to be resolved without delay, one may consider consulting a reference work. The decision to do so reflects the fact that one has a lexicographically relevant need. The need to obtain information about the wingspan of an albatross can be described more generally as a need to acquire knowledge about a specific subject matter.[1] It can be satisfied by means of an appropriate reference work containing relevant data, in this case an encyclopaedia, a bird atlas, or perhaps a dictionary of ornithology.

Offering the user the correct and relevant data is, however, only one of the objectives of lexicography. A great number of reference works do contain the data that satisfy the user's particular need for information, but searching and finding these data may be difficult and time-consuming. This may discourage the user from further consultation of the particular reference work. Indeed, the satisfaction of a need for information depends not only on the presence of specific data, but also on their being accessible in a way that requires the minimum time and effort on the part of the user.

This article investigates access to data in electronic encyclopaedias. The focus is on the satisfaction of the user's need to obtain a precise answer to a specific question rather than to acquire extensive knowledge about a given subject. A brief presentation of access routes available in electronic encyclopaedias and an examination of the notion of access in lexicographical literature reveal that access is customarily conceived of as structures, which may be optimised to allow more effective satisfaction of information needs. This view is challenged here by considering access in terms of a process. The validity of such an approach is subsequently verified by testing three different forms of access to data.

2.   Simple and complex needs for information

The need to acquire or revise knowledge about a specific subject matter has been recognised as lexicographically relevant by scholars representing various theoretical approaches to lexicography. For instance, Tarp (2008) calls it a "cognitive user need" as distinct from a "communicative user need", which is understood as a need for linguistic assistance in an act of communication. This division mirrors the one between addressing general issues ("globale Fragestellungen") and specific issues ("punktuelle Fragestellungen"), proposed much earlier by Hausmann (1977: 144). Both kinds of issues are related here to the area of language, but there is a significant difference between them. Addressing general issues reflects the need to gather factual knowledge, such as "How

is the vocabulary of a particular language structured?" or "What is the lexical stock of language X that deals with the specific subject Y?". In contrast, addressing specific issues reflects the need for a solution to a problem arising in an act of communication, and can be exemplified by such questions as "What is the meaning of X?" or "What is the equivalent of X in language Y?"

From the lexicographical point of view, it does not seem unreasonable to distinguish the need to acquire or revise knowledge from the need to solve communication-related problems. In terms of reference works used to satisfy both kinds of need, the former usually require an encyclopaedia or an encyclopaedic dictionary, while the latter are customarily addressed by various types of general-purpose and specialist dictionaries (although, certainly, both kinds of dictionaries can be used to satisfy cognitive user needs too by offering data that convey information about language or specialised information about a specific topic). More important, however, are the nature and complexity of issues that characterise either kind of need. Looking for a solution to a problem connected with text production, text reception or translation is necessarily confined to linguistic activity and clearly exemplifies a need for specific information. By contrast, gathering knowledge about a given subject is by no means restricted to the area of linguistics and may happen in response to either a general or a specific problem. This assumption and the implication it has for access to data constitute the point of departure for this discussion.

Whatever the area of knowledge one is exploring, one's need for information will fall into one of two broad categories, determined by the degree of complexity of the issue addressed. A simple need is one that can be satisfied by means of a precise, specific, finite amount of information. For example, if one needs to know when the Federal Republic of Germany was proclaimed, a single piece of data will be sufficient: 23 May 1949. In contradistinction to this, if one would like to know how the Reichstag functions, the need for information is complex. It cannot be satisfied by means of a specific piece of data placed in a particular article in a given encyclopaedia. Solving problems of this kind often requires a systematic consultation of one or more reference works.

The distinction between simple and complex needs has a direct bearing on access to data that satisfy them. This is particularly relevant with regard to the needs for information about well-known people, geographical locations, or central concepts within specific domains, such as *function* within mathematics. In many encyclopaedias, articles dealing with such issues contain considerable amounts of data. If one's need for information is simple, e.g. one is only interested to know of what Ferdinand Porsche died, then the prospect of having to peruse all other information about Porsche before one learns what caused his demise might easily put one off using the particular reference work. The relevant data must be reached more efficiently. This confirms that there is a crucial difference between simple and complex needs, but, as will be revealed later in this article, it has not been taken into consideration by metalexicographers.

3.   Access routes in electronic encyclopaedias

Data stored in electronic encyclopaedias (and other reference works designed to help the user with the acquisition of knowledge) can be accessed in a variety of ways. Four basic types of access routes can be distinguished, depending on their starting point, i.e. the place in or outside the given reference works, where the user begins the search for information.

First, data can be searched by means of the list of headwords. All items on the list contain hyperlinks, which, when activated, lead the user directly to the respective articles. This type of access route is available in most encyclopaedias on CD-ROM or DVD-ROM, such as *Britannica 2006 CD-ROM* (2006), as well as in some internet-based encyclopaedias, e.g. *Wikipedia* (2009)[2] and *Aschehougs Leksikon* (2009).

Second, the user can access data via an index of topics or a portal. All items in the index and some items in a portal contain hyperlinks, which, when activated, lead the user to particular articles, either directly or through other, more specific indexes or portals. This type of access route is found in encyclopaedias with a partly thematic access structure, e.g. *Wikipedia* (2009), *Quid* (2009), and *Den Store Danske* (2009). A special type of index of topics is the so-called 'entry menu'. It is an interactive table of contents placed within an article. Obviously, accessing data via entry menus presupposes opening the relevant article in the first place.

Third, data can be accessed through other hyperlinks. These may be placed inside articles, in the outside matter, or outside the given reference work, for instance on an unrelated website or in other software. They can direct the user to particular articles, to the outside matter, or to data placed outside the encyclopaedia. Hyperlinks can be overt, i.e. the format of a particular piece of data (and often the shape of the cursor if placed in the vicinity of the data) immediately make the user aware of the presence of a hyperlink in it. By contrast, covert hyperlinks are invisible to the user. Some encyclopaedias, like *Wikipedia* (2009), offer only overt hyperlinks, others, like *Treccani* (2009), contain only covert hyperlinks,[3] while still others, such as *Larousse* (2009), offer a combination of both kinds.

Fourth, nearly all reference works allow their users to access data through a search box. It is usually possible to narrow the scope of the search by modifying it in a variety of ways. For instance, the search can be limited to a specific part of the encyclopaedia, such as the list of headwords, as in *Wikipedia* (2009), a media collection, as in *Larousse* (2009), an atlas, as in *Uniwersalna Encyklopedia PWN* (2006), or a specific topic, as in *Chronik der Weltgeschichte* (2002). Also, most modern encyclopaedias allow some kinds of Boolean search in order to limit the quantity of data presented as the search result. In very few reference works, such as *Musikordbogen* (2009), the user is additionally obliged to specify the need that has triggered the consultation, e.g. problems with text reception, or the desire to acquire more knowledge about a given topic. Submitting a

query in a search box may yield various results. If the search string matches a headword in the reference work, the respective article is usually opened. Otherwise, a list of relevant headwords is displayed, sometimes with extracts from the respective articles. The extracts normally do not vary according to the search string submitted (exceptions include *Wikipedia* (2009), and other wikis, as well as *Quid* (2009) where the search is powered by Google). The results of searching within other types of data than text may include a list of photographs, films or maps, as in *Larousse* (2009).

The four types of access routes have been analysed and described, both in lexicographical literature and in numerous reference works. Still, very few of them seem to have been devised with a deliberate view to satisfying either simple or complex needs for information. In most cases, the route the user follows terminates in a particular article presented in its entirety, regardless of whether this is necessary or not to satisfy the given need for information efficiently.

The structures provided by the lexicographer neither constitute nor guarantee in themselves rapid and unimpeded access to data if they are not devised with the specific needs and competences of the prospective users in mind. Accordingly, the importance of the relationship between structures and user needs and skills is increasingly recognised by lexicographers (see, e.g. Bergenholtz and Gouws 2007, Gouws 2009). The relevant studies, however, mostly concern dictionaries, often in printed form. Research into needs-adapted access to data in electronic encyclopaedias, especially such that would shed more light on the process taking place during the interaction between the user and the particular structure, has yet to appear.

4.   Research on access to data in electronic reference works

Access to data in electronic reference works has enjoyed wide coverage in literature on lexicography. The research on the subject is characterised by four major features. It is mostly related to dictionaries rather than to encyclopaedias. It is usually done as part of broader studies in electronic lexicography. It seldom explores the relation between access to data and specific types of needs for information. Finally, it often represents an approach which is inadequate for its purpose. Each of these features is examined in more detail below.

Discussion of access to data in electronic reference works usually focuses on dictionaries whose aim is to satisfy the user's need for immediate linguistic assistance. Information concerning data access in electronic encyclopaedias and dictionaries aiming to help the user with knowledge acquisition is much rarer. It usually appears in project reports (Pedersen and Rasmussen 2007, Biffi 2009), reviews (Beißwenger 2002, Eickmeyer 2002, Lobenstein-Reichmann 2002), and comparative analyses of selected reference works (Streitberger 2002).

From the above it can be surmised correctly that access to data is often approached from a practical, descriptive standpoint and usually in connection

with other aspects of electronic lexicography. Theoretical studies whose primary subject is access to data in electronic reference works are rare (though see Bergenholtz and Gouws 2007, Tono 2009, Tarp 2009). To come across theoretical discussions concerning access, one often needs to consult research of a broader scope, such as general descriptions of electronic dictionaries (Nesi 1999, De Schryver 2003) or studies in digitisation of older reference works (Lobenstein-Reichmann 2007).

The scarcity of theoretical studies may explain why few metalexicographers hitherto have shown any interest in the relation between access to data on the one hand and particular types of need for information on the other. Establishing solid theoretical links between these two phenomena belongs to exceptions (Bergenholtz and Gouws 2007, Tarp 2009). Occasionally, one can witness a proposal for some basic principles for a normative theory, as in Nielsen and Mourier (2005). In some other studies, specific types of information needs are not mentioned explicitly but may be deduced from the context. For example, what Beißwenger (2002: 318) surely has in mind when he proposes suggestions for optimising data access for a user who "über das geistige Leben im Hochmittelalter informieren möchte" must be complex needs for information arising in connection with knowledge acquisition. Yet other researchers take the need for information for granted and do not speculate about its type, even though they recognise the importance of securing quick and easy access to data. Thus writes e.g. Streitberger (2002: 99), explaining how users can quench their "kunstgeschichtlicher Wissensdurst" with *Belser Lexikon der Kunst- und Stilgeschichte* (1999). Finally, some metalexicographers reject the issue of adapting data access to particular information needs, as they believe the problem to be insoluble. Trap-Jensen (2005: 115), for example, claims that owing to practical constraints it is simply not possible to devise search options in reference works on the basis of specific types of need for information.

Perhaps the most interesting feature of research on access to data in electronic reference works is the fact that it makes use of theoretical notions which are not quite adequate for its purpose. The definition of the very term *access* is a case in point. As the following examples clearly illustrate, in lexicography, access is commonly perceived in terms of search routes and structures in a reference work, which users follow and enter in order to find data:

> The access structure can be regarded as the search route the dictionary user follows during a dictionary consultation procedure. (Steyn 2004: 276)

> Die Datenakzessivität betrifft die Zugriffsbereitschaft und damit die Nachschlagbarkeit von textuellen lexikographischen Daten. Grundsätzlich gilt: Es muss zwischen der externen und der internen Datenakzessivität unterschieden werden. Weiterhin gilt: Es muss die nichtmediostrukturelle von der mediostrukturellen Datenakzessivität unterschieden werden. (Wiegand 2008: 214)

Such a conception of access represents patterns of cognitive ergonomics seen from the lexicographer's perspective. For the user, it is not only important to

know what search route to follow, but also how to formulate their query, what search result to expect, and, possibly, how much time the entire procedure will take. Moreover, the validity of this approach is mostly confined to printed reference works. For users of electronic lexicographical resources, especially internet-based ones, the mediostructure is not necessarily significant. Data displayed on screen are often retrieved from relational databases, rather than reached via predefined routes through macro-, medio- and microstructures. Certainly, electronic reference works do have access structures, too. However, the preconceptualised search route which underlies the access structure need not be followed by the user during a consultation procedure. The actual search route is frequently through databases and relations between them and is constructed by the user solely for the purpose of a specific act of consultation. Indeed, this is the essence of the electronic revolution, which "lies in the fact that users are liberated from the alphabetical straitjacket, that hypertext, menus, etc. eliminate (artificial) linear text restrictions, that the data conjured up onscreen are not static, and that powerful search capabilities ensure a smooth overarching navigation" (De Schryver 2003: 157). The systems of particular "Leitelemente" (Wiegand 1988) and their mutual relations, devised by the lexicographer, seem to have a diminished significance for the user of an electronic reference work, although it still can be argued that the lexicographer simply does not have to design such relations — they can exploit the possibilities offered by the electronic media.

From the above it appears that *access* is a term denoting a kind of process, where the structure is just one of the elements involved. This view is shared by Bergenholtz and Gouws (2007: 240), who speak of "the notion of an access process […] which includes the access structure as one of its components". This dynamic view is quite convincing, as it can actually be extended so that access does not just denote an instrumental process, but also a constructionist one, defined as follows:

> Access is a process in which the user of a reference work deliberately seeks and finds data which immediately satisfy their specific, lexicographically relevant needs. The process is user-initiated, user-motivated, and user-driven, and it utilizes inner and outer structures of the reference work. It begins with the user's selection of the reference work and their formulation of the query, it makes use of a specific search route, and it terminates once the user has reached the required data, which are found among the search results provided in response to the query.

The process invariably makes use of specific external and/or internal search routes, which are partly or entirely devised by the lexicographer but activated solely by the user. Access is thus observable and thereby amenable to empirical investigation. This can come in the form of both direct observation, including eyetracking analysis, and remote, asynchronous investigation, such as an analysis of log files registering the use of an electronic reference work. The process is also measurable in time, since one can specify its beginning and its

end. The term *access* thereby covers such terms as *search route*, *search string*, *search result* and *search time*.

The validity of the definition proposed above will be tested in the next section. It contains an examination of various forms of accessing data in response to a simple need for information.

5.   Access as a constructionist interaction

To argue in favour of the conception of access as a user-driven process, a short case study is made and presented below. Its aim is to illustrate what happens when the user accesses data in order to satisfy a simple need for information in connection with knowledge acquisition.

5.1   Methodological considerations

The case study consists in testing access to data via three types of access routes which appear to be most appropriate for the retrieval of specific information. These include reaching data through the entry menu placed in the article, through a Boolean search performed in a search box, and through a full text search of HTML text performed in the user's browser. Other types of access routes are disregarded here, each for a different reason. Accessing data solely via a list of headwords does not seem to be particularly suitable for satisfying simple needs for information (cf. the problem of Ferdinand Porsche's death mentioned earlier). Following a hyperlink to a specific article from outside the reference work, e.g. through performing a Google search, does not reflect the conscious choice of that work to satisfy one's need for information. Finally, attempting to reach data through a hyperlink from one article to another is more likely to reflect a stage in satisfying one's complex need for information.

All the tested attempts at accessing data are done in response to a realistic need for specific information.[4] The need can be formulated as the following question: Of what did Frédéric Chopin die? It should be kept in mind, however, that the objective of the test is not to reveal the presence or absence of particular data in a given reference work but to prove certain assumptions about access to data. The data themselves are of secondary importance, as is the truth value of the information they convey.

Only one reference work is subjected to investigation. The choice of *Wikipedia* (2009) for this purpose seems reasonable for several reasons. It is currently the world's biggest general-purpose encyclopaedia for laypeople, it is available in a number of languages, and, being an internet-based reference work, it allows the user to access data via all the routes to be investigated. At the same time it must be borne in mind that *Wikipedia* (2009) is only used here to illustrate the points made. It is not meant as an example to be followed by other electronic encyclopaedias, nor as the most popular or effective reference

work whose use reflects user-initiated search strategies. Users might try to satisfy their information needs more frequently by Google. Answer to quiz-like questions may be sought in specialist reference works, like *Answers.com* (2010). Therefore, wherever applicable, references will be made to other electronic encyclopaedias.

The case study is carried out in controlled conditions. It is a qualitative investigation. There is only one informant involved, and it is the author (hence the frequent use of the first person pronoun in the following sections). The informant is an interested layperson in the area of knowledge to which the information need specified above pertains. Although studies in access to data, which involve only one informant are not uncommon (see e.g. Bergenholtz 2009), there are obvious limitations to what they can reveal. To be of scientific significance, their results must be verified in an experiment involving more informants. The study is thus best regarded as part of a hypothesis formulation, which needs to be tested by a proper experimental procedure.

5.2   Entry menu

Before being able to reach data through an entry menu placed in an article in *Wikipedia* (2009), two other actions need to be performed. First of all, having opened the start page, one must choose the language of the encyclopaedia, and then open the relevant article. I choose the English version. To find out about the cause of Chopin's death, I type "Chopin" in the search box placed in the left menu. Typing a query in the box and clicking the button "Go" will activate a search in the list of articles. Once I type, the type-ahead search is being automatically activated, as a drop-down menu pops up beneath the box, suggesting articles whose headwords begin with the search string. I decide to choose "Chopin". This opens the article, which happens to be of considerable length (slightly over 11 000 words). Only the first section of the article is displayed on screen, the remaining text becoming visible as I scroll down. Not knowing where the data satisfying my need for information are located in the article, to save time I turn to the entry menu. It is rather concise, consisting of twenty items: nine main headings (*Life*, *Memorials*, *Music*, *Works*, *Fiction*, *See also*, *Notes*, *Bibliography*, *External links*) and two sets of altogether eleven subheadings (six under the heading *Life* and five under the heading *Music*) arranged at the second level of nesting (i.e. 1.1.). I click on the subheading *1.6. Final years*, upon which I am redirected to the respective section in the article. The section features two graphic images and a longish text of nearly 900 words. I need to read the entire text, or at least scan it carefully, down to the very last two lines, from which I finally learn that "[i]n 2008, a controversy arose over whether Chopin died of tuberculosis or <u>cystic fibrosis</u> […]." The procedure is rather time-consuming, which makes me wonder whether the same data could be reached more efficiently.

Could the problem possibly have been avoided if the entry menu had been more fine-grained and each item had led to shorter chunks of the article? This cannot be determined with any certainty. The form and content of the entry menu were adequate for my user profile. My prior knowledge about Chopin was restricted to some very basic facts concerning his birth, his death, his nationality, and his love affair with George Sand; my knowledge of his music was and is far from expert. This was nearly mirrored by the simple and concise menu. The fact that it led me to a lengthy text may have dashed my hopes of reaching the data quickly and conveniently, but it did not reveal any problems with the structure of the menu. This was confirmed when I tried to use the French *Wikipedia* (2009) article about Chopin to satisfy the same need for information. The portions of text displayed on activating an item from the entry menu are usually shorter than they are in the English article. However, the menu itself consists of no less than 101 items, spread over five levels of nesting (1.1.1.1.1.). Its usefulness presupposes the user's expert knowledge about Chopin, which I do not happen to have.

The fact that retrieving specific information via an entry menu may not live up to the expectations of some users is perhaps interesting to critics of encyclopaedias. More important, however, is what it reveals about the process of accessing data. The procedure is likely to fail because an interaction between the user and the structure it involves is to a high degree motivated by the lexicographer instead of the user. To begin with, the user has to guess which article to open in order to satisfy their need for information. Whether their guess is correct or not depends in principle on the lexicographer's whims about where to place the relevant data. Next, they need to know which item from the entry menu is most likely to lead them to the required data. This can be much more of a challenge than choosing the correct article. An entry menu invariably represents a prefabricated structuring of information on the basis of some thematic or ontological arrangement of ideas which the lexicographer has related to the notion expressed by the headword. To be of any practical use, the form and contents of the menu must largely correspond to the amount and arrangement of information about the given concept in the user's mind. For obvious reasons, this requirement is unlikely to be met with regard to all intended users of a general-purpose encyclopaedia. Finally, on activating the correct item from the menu, the user would reasonably expect to be redirected to a text in which they can locate the relevant data quickly and conveniently. The length and wording of the text, however, reflect the choices made by the lexicographer, not by the user.

Clearly, placing an entry menu in an article, the lexicographer undoubtedly wishes to help the user retrieve specific information. If rapid and unimpeded data access was just a matter of equipping the reference work with appropriate structures, this solution should always work. That this is not the case confirms that access to data is a process and that any constraints placed on the user's control of it are likely to compromise its success.

5.3   Boolean search

The start page of *Wikipedia* (2009) does not feature a search box which supports Boolean searching in the text of all articles. Such a search box does, however, appear in the left menu when the user has chosen the language of the encyclopaedia. I select English, and, determined to find out quickly and easily what caused Chopin's death, I type "Chopin + died" in the box and click the button "Search". In response to this, the search engine finds all the 783 articles containing both keywords and sorts them by relevance. Extracts from the first twenty articles are displayed on screen. Each one is introduced by its respective headword and contains the key words in context, set in boldface for emphasis. The first article listed among the search results is "Frédéric Chopin". The extract displayed beneath the headword contains occurrences of both "Chopin" and "died", but it can only inform me that Chopin died aged thirty-nine, and that his friend Cherubini had died aged eighty-one in Paris in 1842. This does not satisfy my need for information.

From this point, I can follow two possible routes to reach the relevant data. I can open the article about Frédéric Chopin and look for the data there anyway. Finding them might take time, regardless of whether I choose to read the whole article or use the entry menu instead. It would be of considerable help if all occurrences of the keywords were emphasised inside the article text,[5] but unfortunately this is not the case. Therefore, I decide to abandon the article and examine whether the relevant data appear in the extracts belonging to some other articles. I do not find them in the extracts from the three articles immediately following the first result. However, the fifth result on the list, the article about the role-playing video game *Eternal Sonata*, does contain data satisfying my need for information. The extract informs me that "the game is centered on the Polish romantic pianist and composer Frédéric **Chopin**, who **died** of tuberculosis at the age of 39. [...]". Since I have never heard about the game before, I would not have looked for the answer to my question in the article about it.

The procedure described above provides further incitement to conceive of access to data as a process rather than a structure. It also confirms that the process is user-driven and user-motivated. To satisfy their need for specific information quickly and easily, the user does not have to know which article to open, though, naturally, a certain level of (especially linguistic) competence will facilitate a quick choice. Ideally, they do not have to open any article, as the relevant data are already displayed in one or more extracts featured among the search results.[6] This renders both the composition of the article and the structures the lexicographer has provided to navigate quickly in it totally insignificant.

Admittedly, the user must be lucky in their forming of the search string. While searching for "Chopin + died" in *Wikipedia* (2009) would produce a positive result, submitting "Chopin + death", "Chopin + dead" or "Chopin + demise" would not. The interaction between the user and the structure is only

successful provided the keywords appear in any article in the reference work. On the other hand, even though it does have a bearing on the outcome of the process, the degree of the lexicographer's control of the interaction is in fact very low. The decision concerning the form of the search string resides ultimately with the user.

5.4   Full text search of the HTML text

Since *Wikipedia* (2009) is internet-based, and all its articles are displayed as HTML texts, there exists yet another way of quickly and easily retrieving specific information from it. On the start page of the encyclopaedia, I choose the language (English) and then open the article "Frédéric Chopin", hoping that it features the data which will satisfy my need for information, which is to learn what caused Chopin's death. Once the article is displayed on screen, I hit two keys on the keyboard: *Ctrl* and *F*. This activates the search window in my Internet browser (Internet Explorer). Into it I type "died", to which the browser replies by highlighting all occurrences of the search string in the article text.[7] The first occurrence is found towards the end of the lengthy introductory section of the article in the following context: "Always in frail health, he **died** in Paris in 1849, aged thirty-nine, of <u>pulmonary tuberculosis</u>." It happens to satisfy my need for information. If this had not been the case, I would have examined the next occurrence of "died", to which I would have been redirected by clicking on the button "Next" in the browser's search window.

Certainly, the success of this form of data access depends on the correct choice of the article which features the relevant data, and this, in turn, has been determined by the lexicographer. However, once the article is displayed on screen, the user's reliance on the structures provided by the lexicographer diminishes. Actually, the structures are abandoned in favour of a solution provided by the programmer of the Internet browser. The lexicographer certainly still deserves some of the credit if they have incorporated the solution deliberately with a view to facilitating the access process for the user. On the other hand, if this is the case, one may wonder why performing text search of the HTML document is not mentioned in user's guides to online encyclopaedias,[8] together with other instructions on how to search for information. If the user is not informed of this possibility, chances are that they successfully construct their own access process by exploiting the medium of the reference work regardless of whether that was the intention of the lexicographer. Such cases cannot be accounted for if access to data is discussed solely in terms of lexicographical structures.

6.   Final remarks

The objective of this article was to provide a new insight into access to data in electronic reference works. It was argued that access should not be viewed as a

structural, instrumental notion determined by the lexicographer, but as a process initiated, constructed and controlled by the user.

The efficiency of this process depends on two factors. First, access to data needs to be adapted to the specific type of the user's need for information. The examination of three forms of access to data carried out in the article was restricted to the issue of satisfying simple needs for information. If the user's need had been a complex one, the process of accessing data would have been different. This matter requires advanced, experimental research into how users precisely behave when they look for data in specific settings. Search behaviour must be regarded as a cognitive interaction, and future research must include psychological manifestation of this cognitive activity. This has been revealed, although to a very modest extent, in the case study. Second, reference works must be constructed in a way that allows the user maximum control of the process of accessing data, including adaptive design.

The reference works referred to in the article were general-purpose encyclopaedias for laypeople. It should be kept in mind that access to data in connection with the acquisition of knowledge is an area of research that should not be limited to the genre investigated. Satisfaction of the same lexicographically relevant needs may be sought in other reference works, such as specialised dictionaries and handbooks, which account for an overwhelming majority of reference works produced, but still attract relatively little attention of metalexicographers. Investigating such works may reveal more insights into the relation between access to data, information needs of the user, and particular lexicographical genres aiming to satisfy those needs.

Endnotes

1. Although a layperson might experience the need for this information, the exact question asked is likely to differ from the one presented here. As *wingspan* might not be part of the vocabulary of such user, a more probable form of the question is "How big is an albatross?" Here, however, for the sake of precision the question is formulated in such a way as to clearly indicate the precise need for information.

2. Admittedly, this access route is not the one most commonly used in *Wikipedia* (2009).

3. In *Treccani* (2009) the user is actually made aware of the fact that all words in the article contain hyperlinks. A bar featuring this information always appears above the displayed article.

4. This excludes information needs based on false presuppositions, such as the need to know the name of Nicolas Sarkozy's sister (he has no sister), and derived from patently self-contradictory premises, such as the need to know the place of death of the Sultan of the Vatican Republic who ruled in the period 3045–3010 AD.

5. This feature is found in e.g. *Aschehougs Leksikon* (2009) and *Den Store Danske* (2009).

6. This requires, however, that the extracts always vary according to the search string submitted.

7. This applies also to Google Chrome and Mozilla Firefox.

8.   The only counterexample known to the author is that of the former online version of the Swedish *Nationalencyklopedin*. The user's guide to the current version does not contain the information on accessing data via full text search of an article.

References

Reference works

Answers.com. The world's leading Q&A site. http://www.answers.com [accessed 4 August 2010].

Aschehougs Leksikon. http://www.aschehougsleksikon.dk [accessed 4 December 2009].

Belser Lexikon der Kunst- und Stilgeschichte. Version 2.0. Munich: USM.de Systhemor, 1999.

Bergenholtz, I. 2009. *Musikordbogen*. Århus: Handelshøjskolen i Århus. http://www.musikord-bogen.dk [accessed 4 December 2009].

Britannica 2006 CD-ROM = Encyclopaedia Britannica 2006 Deluxe Edition CD-ROM. London: Encyclo-paedia Britannica Inc., 2006.

Chronik der Weltgeschichte Version 3.0. Der neue Kulturfahrplan nach Werner Stein. CD-ROM. Munich: USM.de Systhema, 2002.

Den Store Danske. Gyldendals åbne encyklopædi. http://www.denstoredanske.dk [accessed 4 December 2009].

Larousse = Encyclopédie contributive Larousse.fr. http://www.larousse.fr/encyclopedie [accessed 4 December 2009].

Nationalencyclopedin. http://www.ne.se [accessed 5 March 2010].

Quid = L'Encyclopédie Quid. http://www.quid.fr [accessed 4 December 2009].

Treccani — Enciclopedia e Vocabolario on line, Articoli, Video. http://www.treccani.it [accessed 4 December 2009].

Uniwersalna Encyklopedia PWN = Uniwersalna encyklopedia multimedialna PWN. (Wersja 1.) Warsaw: PWN, 2006.

Wikipedia. *The Free Encyclopedia*. http://www.wikipedia.org [accessed 4 December 2009].

Other literature

Beißwenger, M. 2002. (Review of) Chronik der Weltgeschichte Version 3.0. *Lexicographica* 18: 312-320.

Bergenholtz, H. 2009. Hurtig og sikker tilgang til informationer om ordforbindelser. *LexicoNordica* 16: 29-54.

Bergenholtz, H. and R.H. Gouws. 2007. The Access Process in Dictionaries for Fixed Expressions. *Lexicographica* 23: 237-260.

Biffi, M. 2009. Accademia della Crusca's Online Dictionaries. Bruti, S., R. Cella and M.F. Albert (Eds.). 2009. *Perspectives on Lexicography in Italy and Europe*: 239-286. Newcastle upon Tyne: Cambridge Scholars Press.

De Schryver, G.-M. 2003. Lexicographers' Dreams in the Electronic-dictionary Age. *International Journal of Lexicography* 16(2): 143-199.

Eickmeyer, J. 2002. (Review of) Reclams elektronisches Romanlexikon. Deutschsprachige erzählen-de Literatur vom Mittelalter bis zum Gegenwart. *Lexicographica* 18: 311-312.

Gouws, R.H. 2009. Sinuous Lemma Files in Printed Dictionaries. Access and Lexicographic Functions. Nielsen, S. and S. Tarp (Eds.). 2009. *Lexicography in the 21st Century. In Honour of Henning Bergenholtz*: 3-21. Amsterdam/Philadelphia: John Benjamins.

Hausmann, F.J. 1977. *Einführung in die Benutzung der neufranzösischen Wörterbücher.* Romanistische Arbeitshefte 19. Tübingen: Max Niemeyer.

Lobenstein-Reichmann, A. 2002. (Review of) Lexikon des Mittelalters. *Lexicographica* 18: 300-301.

Lobenstein-Reichmann, A. 2007. Allgemeine Überlegungen zur Retrodigitalisierung historischer Wörterbücher des Deutschen. *Lexicographica* 23: 173-198.

Nesi, H. 1999. A User's Guide to Electronic Dictionaries for Language Learners. *International Journal of Lexicography* 12(1): 55-66.

Nielsen, S. and L. Mourier. 2005. Internet Accounting Dictionaries: Present Solutions and Future Opportunities. *Hermes* 34: 83-116.

Pedersen, J.H. and T.E. Rasmussen. 2007. Strategi i navigation. *LexicoNordica* 14: 71-80.

Steyn, M. 2004. The Access Structure in Learner's Dictionaries. *Lexikos* 14: 275-298.

Streitberger, A. 2002. Wörterbücher der Kunstgeschichte. *Lexicographica* 18: 89-103.

Tarp, S. 2008. *Lexicography in the Borderland between Knowledge and Non-Knowledge. General Lexicographical Theory with Particular Focus on Learner's Lexicography*. Lexicographica. Series Maior 134. Tübingen: Max Niemeyer.

Tarp, S. 2009. Reflections on Data Access in Lexicographic Works. Nielsen, S. and S. Tarp (Eds.). 2009. *Lexicography in the 21st Century. In Honour of Henning Bergenholtz*: 43-61. Amsterdam/ Philadelphia: John Benjamins.

Tono, Y. 2009. Pocket Electronic Dictionaries in Japan: User Perspectives. Bergenholtz, H., S. Tarp, and S. Nielsen (Eds.). 2009. *Lexicography at a Crossroads. Dictionaries and Encyclopedias Today, Lexicographical Tools Tomorrow*: 33-67. Berlin: Peter Lang.

Trap-Jensen, L. 2005. Virtuelle perspektiver for ordbogsredigering: muligheder, strategier og virkelighedens begrænsning. *LexicoNordica* 12: 109-121.

Wiegand, H.E. 1988. Wörterbuchartikel als Text. Harras, G. (Ed.). 1988. *Das Wörterbuch. Artikel und Verweisstrukturen. Jahrbuch 1987 des Instituts für deutsche Sprache*: 30-120. Sprache der Gegenwart LXXIV. Düsseldorf: Schwann.

Wiegand, H.E. 2008. Zugriffsstrukturen in Printwörterbüchern. Ein zusammenfassender Beitrag zu einem zentralen Ausschnitt einer Theorie der Wörterbuchform. *Lexicographica* 24: 209-316.