The Construction of Online Health TermFinder and its English–Chinese Bilingualization

Jun Ding[1], Pam Peters[2], Adam Smith[2]

[1] Fudan University, No 220 Handan Rd. Shanghai, China.
[2] Macquarie University, NSW 2109, Australia.
E-mail: jding@fudan.edu.cn, pam.peters@mq.edu.au, adam.smith@mq.edu.au

Abstract

Health TermFinder (HTF) is an online platform and information tool designed to support medical and health terminologies. Pilot termbanks in selected fields such as breast cancer are currently under construction at Macquarie University in Sydney. Cooperation with Fudan University in Shanghai is underway to develop a bilingualized English–Chinese version of HTF. This paper provides a theoretical overview of HTF as a customized electronic information tool, with reflections on its structure, data organization, user interface and overall principles of construction. Following a discussion of the macrostructure of HTF, i.e., whether it is essentially a lexicographic or terminological work, two sections of the paper are devoted to discussions of its corpus-based selection of headwords and design of the microstructure, with emphasis on the user-oriented philosophy underlying both and based on best principles/practice in lexicography and multimodal language learning. The status quo of the cooperative bilingualization project is given close examination in Section 5, and in Section 6 the possible use of adaptive hypermedia in its future development is proposed.

Keywords: Online dictionary; Health TermFinder; user-oriented; bilingualized; adaptive hypermedia

1. Introduction

The difficulties and problems arising from the use of medical terminology cannot be overestimated in either medical research or in practice. The high linguistic demands of the language found in online health information, which could cause problems for those with low levels of literacy in English, motivated researchers at Macquarie University, Sydney1, to construct a public online information tool for medical terminologies, codenamed Health TermFinder (HTF). Its target users include second-language health professionals in Australia and native English speakers without tertiary education. The Macquarie team is currently working on the first of the HTF termbanks consisting of breast cancer terminology, which currently comprises 51 pages.

[1] This team includes the two coauthors for this paper, Pam Peters, director of the TermFinder project and Adam Smith, researcher. Others are lexicographer Yusmin Funk, and Professor John Boyages of the Macquarie University Cancer Institute, who reviews the termbank's medical content for accuracy.

Meanwhile, the cooperative project of bilingualizing HTF into Chinese at Fudan University, Shanghai, is under negotiation with a team of English–Chinese bilingual lexicographers. The bilingualized Online Health TermFinder (BHTF) is expected to meet the needs of medical students at the Medical School of Fudan University (both undergraduates and graduate students) at its initial stage of development. Once in its later and more full-fledged form, BHTF will be made accessible to the whole Mandarin-speaking community in China.

So what is the nature of this Online Health TermFinder? Is it essentially a lexicographical or a terminological work? If, as described above, the project seems to have begun with observations on specific needs of specific sets of users, upon which principles is its design based; what are its macro- and micro-structures? And what makes the English–Chinese bilingualized version special in comparison to the plain translations into other Australian community languages (including Chinese) offered on the HTF platform? These are the questions to be addressed in this paper which attempts to examine not only the design and input data, but also the construction philosophy of HTF.

2. Lexicographic or Terminological?

In a broad sense, HTF is designed to be an online dictionary-type tool, providing help with health-related and medical terms in English. Yet initially it follows the so-called onomasiological model: a certain health issue is selected as the subject field for the new termbank. For instance, HTF currently includes only one such specialized area, the breast cancer termbank. However, the contents of the termbank do not represent a structured vocabulary of terms used in the field, nor are they restricted to concepts related only to breast cancer. HTF termbanks deal with not only medical terminologies, but also semi-technical terms. This is because their target users are people with low literacy levels in English, including both second-language health professionals and native-English-speaking patients and carers without tertiary education. Since semi-technical terms are usually inherently polysemous, they are likely to pose difficulties to the target users. Terms such as *treatment* will be searchable from one termbank to another, as many are generic medical terms useful to people with different medical problems. Therefore, despite its essentially onomasiological structure (consisting of distinct medical fields), HTF could hardly be considered a strictly terminological project (Riggs, 1989: 89) in view of the mixed lexical content of individual termbanks. Moreover, HTF is designed to serve decoding, or interpretive, purposes at the functional level; another reason to categorize it as essentially a lexicographic rather than terminological work, since the latter is usually also defined by its aim "to help writers produce texts" (Riggs, 1989: 90).

Though lexicographic by nature, HTF also differs considerably from a medical dictionary. For one thing, it lacks the scale or all-inclusiveness of a standard print dictionary. Unlike many online specialized dictionaries, it does not have a printed

counterpart. In other words, it is not adapted from a medical dictionary already in existence. The entry terms included in the breast cancer termbank are instead extracted from a database of online documents on breast cancer care, built by the team at Macquarie University. This practice of building reference databases from scratch will be replicated for other fields of healthcare. Based on such databases, HTF will eventually develop into a huge online multidisciplinary clearing-house in healthcare, rather than a conventional medical dictionary.

This also means that each individual termbank will have a claim to independence, and thus can be made available to users as a stand-alone termbank. In other words, it is not be necessary to wait until the whole project is completed before launching it for public use, unlike the case of most dictionaries which have to be finished from A to Z before going into print or online. The HTF project ought thus to be looked upon as a process rather than a product. Its construction would simply go on until all the important health and medical areas are dealt with, and after that it could still be maintained in a continuously updatable form. Since users' needs are not static, but change and develop throughout time, the updatable form of HTF makes it a lexicographic work which can be constantly adapted and modified to meet the new or evolving needs of its users.

3. User-oriented data

In his discussion of lexicography for the language learner, Tarp (2008) elaborated on the importance of knowing the user profile, user situation, and user needs when creating an online dictionary tool. HTF is exactly such a lexicographic work, designed with a clear extra-lexicographic identification of its specific set of users and their specific needs.

The problems caused by medical terminology are a constant challenge for those health professionals in Australia who speak English as a second language. Native speakers with low levels of literacy encounter similar difficulties in understanding the "jargon" of medicine when either communicating with their doctors or reading printed factsheets or medical websites to access more in-depth information. Researchers at Macquarie University were thus motivated to construct an online information tool for medical terms so as to provide post-consultation help to patients and carers, as well as linguistic support to second-language health professionals.

A large body of online documents on breast cancer were collected from one of Macquarie University Library's specialized online LibGuides[2]. They were categorized into two types in view of different readerships: those designed for the general public and those for health professionals. The documents were accordingly extracted into two separate databases: public (521,232 words) and professional (514,830 words, as of December 2014). Contents of the documents and their respective target audience are

[2] http://libguides.mq.edu.au/content.php?pid=379776&sid=3605261

listed in Appendix 1. Data analyses were carried out by the Macquarie research team to extract word frequencies and other lexical statistics (Peters et al., 2015)[3]. A preliminary table listing the top 24 words and terms in the professional and public databases are presented in Appendix 2. The very high levels of medical and semi-technical health management terms (*clinical, biopsy, carcinoma, screening*) in the professional listing show the demands on second-language professionals, let alone lay readers (patients and their carers) with low literacy levels in English.

All this preliminary terminological research demonstrates the user-oriented philosophy for HTF. The two databases are also being used as a corpus for the compilation of the breast cancer termbank; namely, for identifying terms, prioritizing them for attention, and providing examples of their usage in technical texts. Since the medical documents in the corpus are up-to-date and have specific readerships (breast cancer professionals and patients), the data extracted from them are highly user-oriented and consequently ensure the uptodateness and usefulness of the definitions and examples for the headwords entered in the termbank.

4. User-oriented Microstructure

The microstructure of each head entry is based on best practice for learners' dictionaries as well as multimodal language learning (Lemke, 1998). The users' actual needs are not easily ascertained through questionnaires or interviews, since "users may frequently only have a vague or approximate idea of the objective needs" (Tarp, 2009: 281). On the other hand, profiling the vocabularies of professional vs. public online documents on breast cancer is a practical and productive way of discovering the "genuine or objective needs occurring before the consultation process, i.e., extra-lexicographically" (Tarp, 2009: 282). That the needs thus identified are hypothetical does not make them invalid; though of course the validity still needs to be assessed by users once they start using the HTF termbanks.

Because HTF is a nonprofit research project, freely available to the public, no consideration would be given to the artificial needs of potential users, which are defined by Tarp (2009: 282) as publicity-created subjective needs mainly of interest to commercial publishers. Instead, the content and arrangement of information on each HTF page is designed to meet the genuine, objective needs of the target users. Below is a screenshot of the breast cancer termbank page for the word "lymphoedema", showing the essential English content.

[3] This article is titled "Language, Terminology and the Readability of Online Cancer Information".

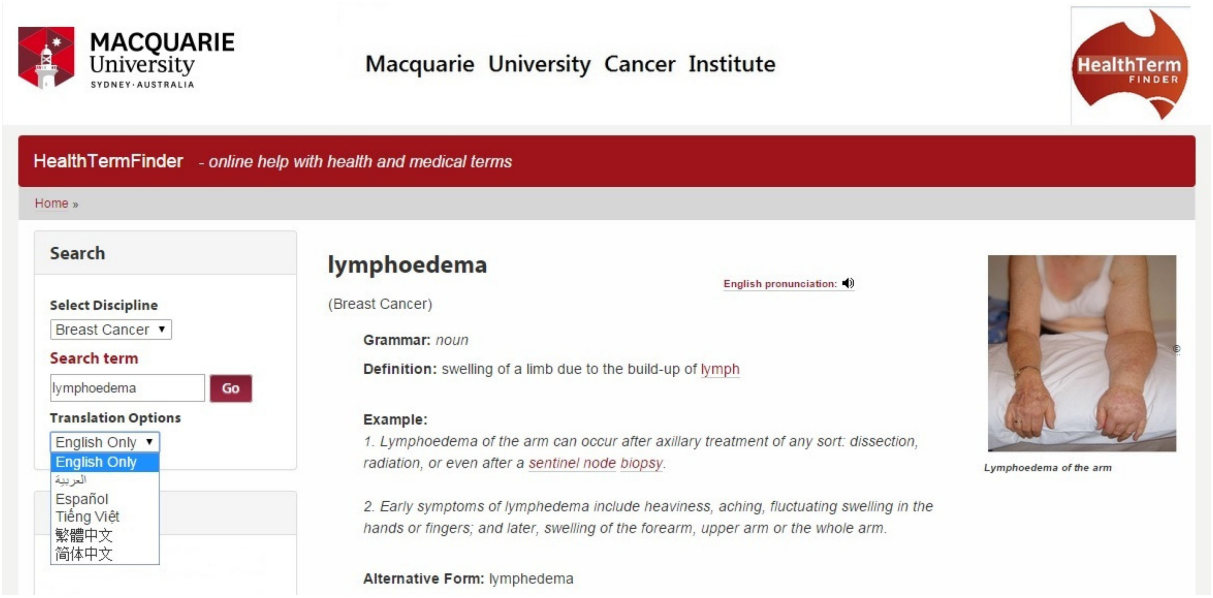

Figure 1: Screenshot of the page for "lymphoedema"

As we can see from the Figure, each term page in HTF includes five elements of lexicographic information:

1) lemma: *lymphoedema*

2) grammatical label: *noun*

3) definition: *swelling of a limb due to the build-up of lymph*

4) examples: *1. Lymphoedema of the arm can occur after axillary treatment of any sort: dissection, radiation, or even after a sentinel node biopsy. 2. Early symptoms of lymphedema include heaviness, aching, fluctuating swelling in the hands or fingers; and later, swelling of the forearm, upper arm or the whole arm*

5) alternative form: *lymphedema*

One of the foremost features of HTF is that the definitions are drafted in plain, highly-accessible English, accordant with the needs of second-language users and those with low reading skills. The definitions are induced from actual instances of the term's use in the corpus, to cover both intensive and functional aspects of its meaning as far as possible.

Also noteworthy is the fact that neither captions nor labels on HTF are given in abbreviations or initials. It is common practice in dictionary making to avoid using captions (such as "Definition", "Examples") and to present grammatical labels as briefly as possible ("n" for "noun", for instance), so as to save precious space for more indispensable information. Space is no longer a problem with online information tools and given that the main concern of HTF is to make the look-up process as easy and

friendly as possible for its target users, we have retained captions and labels spelled out in full to serve as important signposts.

For each entry term two examples of usage are selected from the corpus to complement the definition and provide users with both linguistic and factual information for the term in question. Illustrative materials are also sought in the corpus to show the term's place among other related terms, usually arranged in labeled diagrams or tables of parallel terms. Diagrams, tables, and pictures of relevant images, such as the above one showing "lymphoedema of the arm", are introduced based on Lemke's (1998) theories about meaning-making via various semiotic "channels". Lemke purports that information passed through different channels, such as linguistic, visual, pictorial and acoustic, can be equivalent or complementary, and may or may not reach the person simultaneously. Multimedia facilities make it possible to incorporate multiple semiotic systems on HTF. Besides the visual presentations of graphs, tables, pictures, etc., audio files providing the pronunciation and definition of the term are also available on each term page.

On the left-hand part of the page one can select the relevant termbank (breast cancer, for instance), and can then search terms. Below the look-up box, translations are offered in four of the major community languages in Australia, namely Arabic, Spanish, Vietnamese and Chinese (both traditional and simplified), which, when selected, raise translation boxes for the head term and its definition, as well as for the captions on graphics and labels on diagrams. The translated elements are expected to provide the second-language users with a more efficient "channel" for accessing the relevant information and anchoring their understanding. All the primary contents of HTF (definitions, examples, images, tables) are reviewed by medical experts, and "checkers" are appointed to review the primary translations for each language.

5. English–Chinese Bilingualization

Though translations of the primary contents are available in four languages for selected elements on HTF pages, the system will be fully bilingualized into Chinese (simplified) only in the second stage of the project, which will be carried out at the English department of Fudan University, Shanghai. Again, the English–Chinese bilingualization project is based on a clear identification of target users and their needs. The plan is to make the bilingualized Health TermFinder (BHTF) first accessible to Chinese students of the Medical School at Fudan for training purposes. With its medical terms related to different specific diseases, BHTF can be used as a specialized reference tool alongside more general English–Chinese medical dictionaries.

Ever since Benjamin Hobson (1816–1873) published *A Medical Vocabulary in English and Chinese* (1858), the earliest English–Chinese medical glossary of its kind known in China, the translation of medical terminologies from English into Chinese has played a role of pivotal importance in the development of medical science in the country (Wu &

Wong, 1932; Chen, 1984; Fu, 1990; Ma et al., 1993; Sun, 2010). One hundred and fifty years later, most important medical terms now have Chinese equivalents well established in the language (for instance, 乳腺癌[ruxian' ai] for *breast cancer*, 淋巴[linba] for *lymph*, 血管[xueguan] for *blood vessel*, etc.). For obvious reasons, medical and health professionals in China have to learn English and conduct their research and practice medicine using English as a second language. As a result, medical dictionaries are always in great demand and the best ones are often based on authoritative monolingual English medical dictionaries. For instance, *An English–Chinese Medical Dictionary* (ECMD, editor-in-chief Weiyi Chen, 1984, 1997, 2009, 2014) was largely a translation work of *Dorland's Illustrated Medical Dictionary* (Li & Chen, 2006). These bilingual medical dictionaries are also being increasingly converted into digital forms. The 3rd edition of ECMD was developed into a mobile phone application and is available for free downloading. Yet, the majority of these dictionaries offer only Chinese equivalents. No definitions for the head entries either in English or Chinese are included.

BHTF aims to provide Chinese medical students with more complete information about medical terms, including all the English texts for the head entry, its Chinese equivalent(s), Chinese translations of the definitions and examples and also of all English terms and texts in the diagram/table/illustration and usage notes. Although the Chinese equivalents of the English definitions would be sufficient for Chinese medical students to understand the looked-up term, their constant need to improve their level of English would drive them to read the English texts. In fact, BHTF can also serve as an alternative language learning tool for such users. Meanwhile, it is also necessary to provide Chinese translations of the definitions because of the students' limited English proficiency. Moreover, the independently-drafted definitions could also deviate from the orthodox ones (i.e. those found in traditional medical dictionaries) because the rapid development in medical science may impose some newly acquired, context-specific meanings on established terms. Definitions in Chinese could thus more efficiently alert the users to these differences. It is common practice for bilingual dictionaries in China to present translations of all illustrative examples, and would therefore be expected by Chinese-speaking users and would aid their comprehension of the head term and their English learning in general.

At a later stage, BHTF is to be made available to the general Chinese-speaking public, who often need help with medical terms after consulting a health professional. This stage will occur after the Australian HTF is made bidirectional, i.e., equipped with a redeveloped version of the present platform as a Chinese–English structure. Medical terms in both English and Chinese will then be searchable on the platform, each navigating users to the same entry page for information. Users from the Chinese public, with a limited command of English, are likely to benefit most from the Chinese translations for the looked-up term. However, if able to look up Chinese terms on BHTF, this online information tool will be doubly useful, providing a tool for Chinese citizens as well as for medical students. Also under consideration is the Romanization

of the Chinese equivalents, i.e., the inclusion of their Pinyin forms. This is because an increasing number of foreign students are coming to study at Fudan University each year. These non-Chinese-speaking students may need to communicate about medical issues in Chinese, and the Romanization of Chinese characters would considerably facilitate their pronunciation (the pronunciation of the Chinese characters cannot be inferred from their form). Audio files of Chinese equivalents and definitions can also be provided for their benefit as well as for that of Chinese citizens who speak a regional dialect.

6. Adaptive Hypermedia – Future Development for BHTF

The Fudan team working on the BHTF are considering the application of hypermedia to the termbank as part of its future development. *Hypermedia* in this context refers to user-adaptive software systems which can select and prioritize items of information for users depending on their individual needs (Brusilovsky & Millan, 2007). Adaptive hypermedia has been applied to an English dictionary of finance for Indonesian students (Kwary, 2011), in which the adaptive search system directs the user's search action to the results decided by the system to be preferable or most suited to the user's needs. It requires lexicographers to decide upon the most suitable result when searching for a particular term, and to set it up accordingly in the dictionary system. However, these decisions must be based on the user profile.

Since we have a very specific group of target users for the first stage of BHTF – Chinese medical students at Fudan University – it would be relatively easy to build a comprehensive user profile. Then, for example, we would be able to decide if, for a certain medical term, its Chinese translations would be more helpful to the Chinese student than would the English definition, or vice versa. As medical students, these users are expected to be equipped with a greater knowledge of terms than non-professionals, so that when they look up a certain English medical term, it is likely to be because they want to read its definitions in English and to see examples illustrating its actual usages. In other words, medical students are very likely to use BHTF for productive purposes as well as receptive ones, though the original English version is designed for meeting decoding needs. If a semi-technical term is looked up, it may suggest that the student's level of scientific English is below average, and therefore it is best to direct the user immediately to the Chinese translations which can solve their decoding problems more efficiently.

This kind of hypothesizing is based on what Tarp calls "function-related needs"; needs identified as objective and in an extra-lexicographic situation, and differing from "usage-related needs" which occur only during the actual consultation process (2009: 283). For instance, it may happen that when a certain esoteric medical terminology is looked up by a user and they are offered its English definition and examples, they always move straight on to its Chinese translations. This could imply that the term in question is completely new to most medical students who look it up on BHTF, and

that its English definitions may not be clear enough for them after all. It is equally possible that a semi-technical term is looked up more often for its English than its Chinese parts, which could mean that this term is familiar to most users and more likely to pose difficulty in encoding rather than decoding tasks. In that case, what is required is another adaptive system called "log file action analysis" (Kwary, 2011: 37) which saves the users' search actions for different terms. The initial setup for that particular term may be automatically changed after a number of such actions being recorded.

The second stage of BHTF would present a more complex scenario, when is the resource is made accessible to the general community, composed of mostly Chinese citizens with on average a very limited command of English. Yet, over the years, scholars interested in future dictionaries have discussed and predicted the possibility of individualization or customization of dictionaries (Dodd, 1989; Atkins, 1996; Whitelock & Edmonds, 2000; de Schryver, 2003; Lew & de Schryver, 2014); each discussion more daring and confident than the previous. Indeed, given the speed and scale of technological development, one has every reason to feel confident about the advent of ever more advanced adaptive hypermedia software in the near future which could save and process search tasks performed by individual IP addresses and adapt the system to any particular user's needs.

7. Conclusion

Online Health TermFinder, which is currently under construction at Macquarie University, is a completely user-oriented, nonprofit, digitalized lexicographical project aimed at providing linguistic and factual information on medical terms with open access on the internet. HTF targets users are either health professionals speaking English as a second language or native-English speakers with low literacy levels, and serves predominantly decoding purposes. Its bilingualized version, BHTF, to be constructed at Fudan University, will primarily target Chinese medical students with both decoding and encoding needs.

With the current development focused on the field of breast cancer, directions for expansion are already under consideration within the Macquarie research team. Other types of cancer and major medical and health problems such as orthopedics and mental health also call for online information from well-designed termbanks. Meanwhile, after consulting the Medical Faculty, the Fudan BHTF team has nominated priority areas to align with the structure of medical training, which include the field of various cancers, respiratory diseases, and paediatrics. Once a new area for development has been identified, online materials in English in the relevant field will be sought and collected to build databases and thus a new round of TermFinder construction will begin.

8. References

Atkins, B. T. S. (1996). Bilingual Dictionaries: Past, Present and Future. In M. Gellerstam et al. (eds.) *Euralex '96 Proceedings I-II, Papers submitted to the Seventh EURALEX International Congress on Lexicography in Goteborg, Sweden.* Gothenburg: Department of Swedish, Goteborg University, 515-546.

Brusilovsky, P. & Millan, E. (2007). User Models for Adaptive Hypermedia and Adaptive Educational Systems. In P. Brusilovsky, A. Kobsa & W. Nejdl (eds.) *The Adaptive Web.* Berlin: Springer-Verlag, 3-53.

Chen, Bangxian. (1984). *History of Chinese Medicine.* Shanghai: Shanghai Bookstore.

Dodd, W. S. (1989). Lexicomputing and the Dictionary of the Future. In G. James (ed.), *Lexicographers and Their Works.* (*Exeter Linguistic Studies 14.*) Exeter: Exeter University Press, 83-89.

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

Fu, Weikang. (1990). *History of Chinese Medicine.* Shanghai: Shanghai Chinese Medicine University Press.

Kwary, D. A. (2011). Adaptive Hypermedia and User-Oriented Data for Online Dictionaries: A Case Study on an English Dictionary of Finance for Indonesian Students. *International Journal of Lexicography*, 25 (1), pp. 30-49.

Lemke, J. (1998). Multiplying Meaning: Visual and Verbal Semiotics. In J. Martin & R. Veel (eds.) *Reading Science.* London: Routledge. Available at: http://academic.brooklyn.cuny.edu/education/jlemke/papers/mxm-syd.htm

Lew, R. & de Schryver, G.-M. (2014). Dictionary Users in the Digital Revolution. *International Journal of Lexicography,* 27(4), pp. 341-359.

Li, D. & Chen, W. (2006). The English-Chinese Translation of Medical Terms. *Chinese Translators Journal*, 6, pp. 60-64.

Ma, Boying, Gao, Xi & Hong, Zhongli. (1993). *History of Medical Culture Communication between China and the World.* Shanghai: Wenhui Publish House.

Riggs, F. W. (1989). Terminology and Lexicography: Their Complementarity. *International Journal of Lexicography*, 2 (2), pp. 89-110.

Sun, Zhuo. (2010). The Creation of Modern Medical Terms: A Case Study of Benjamin Hobson and His A Medical Vocabulary in English and Chinese. *Studies of Natural Science History.* vol. 4. Available at: http://www.cssn.cn/st/st_xsplc/201404/t20140410_1063151.shtml

Tarp, S. (2008). *Lexicography in the Borderland between Knowledge and Non-knowledge: General Lexicographical Theory with Particular Focus on Leaner's Lexicography.* Niemeyer: Tübingen.

Tarp, S. (2009). Reflections on Lexicographical User Research. *Lexicos*, 19, pp. 275-296.

Whitelock, P. & Edmonds, P. (2000). The Sharp Intelligent Dictionary. In U. Heid et al. (eds.), *Proceedings of the Euralex International Congress, EURALEX 2000, Stuttgart, Germany, August 8th-12th, 2000.* Stuttgart: Institut fur Maschinelle Sprachverarbeitung, Universitat Stuttgart, pp. 871-876.

Wu, L. T., & Wong, C. (1932). *History of Chinese Medicine*. Tientsin: Tientsin Press. Ltd.

Appendix 1: Contents of the databases

document	*words*	*target audience*
Cancer Council NSW	117738	public
Cancer Council Australia	10265	public
National Cancer Prevention and Early Detection Policy	11270	public
Cancer Council Victoria brochures	37993	public
Cancer Australia website	87993	public
BC in men	6556	(men) public
Clinical best practice and info for health professionals	497760	professional
Cancer Australia pamphlets	66830	public
Breast cancer risk factors: a review of the evidence 2009	38899	professional
All BCI pamphlets in word doc	58308	public
Breast Cancer network Australia website	166486	public
Information for health professionals	876	professional
BCNA pamphlets	114102	public
National Breast Cancer Foundation_part of website	3309	public
ABC Health & Wellbeing - Breast Cancer	22494	public
Pink Hope	16863	public
pink hope pamphlets	16630	public
Life After early Breast Cancer	20606	public
Breast Cancer and Axillary Lymph Nodes	644	public
BRCA Genes and Breast Cancer	622	public
TOTAL	1296244	

Appendix 2: Comparative rankings of top 24 words and terms in the two databases

Professional		514830 wds		**Public**		521232 wds
rank	term	frequency		rank	term	frequency
1	breast	11204		1	cancer	10730
2	cancer	10403		2	breast	9565
3	women	4724		3	women	5167
4	risk	2437		4	treatment	2932
5	clinical	2424		5	information	2011
6	treatment	2017		6	risk	1863
7	patients	1687		7	help	1550
8	study	1492		8	care	1447
9	evidence	1387		9	health	1447
10	practice	1286		10	surgery	1275
11	management	1269		11	people	1172
12	information	1227		12	reconstruction	1108
13	biopsy	1188		13	support	1101
14	guidelines	1138		14	time	1083
15	imaging	1137		15	research	1034
16	national	1116		16	pain	1012
17	Australia	1111		17	family	975
18	carcinoma	1067		18	chemotherapy	963
19	diagnosis	1050		19	find	931
20	care	1027		20	Australia	917
21	early	1007		21	feel	914
22	studies	991		22	side	811
23	health	937		23	effects	800
24	screening	924		24	doctor	790