Dictionary Writing Systems and Corpus Query Systems

Survey short title: DWS and CQS

Survey long title: Dictionary Writing Systems and
Corpus Query Systems

Question number: 94

Survey is active

Active from: 23.08.2014   Active until: 20.12.2014

Author: krek.simon@gmail.com   Edited: krek.simon@gmail.com

Date: 24.07.2014   Date: 19.11.2014

Q1 - Please provide your name(s) and surname(s) - separated by commas if more responders are providing answers together:

Q2 - Please provide the name of the country where you work:

Q3 - Please provide the name(s) of the institution(s) where you work (e.g. Institute of the Estonian language, Jozef Stefan institute etc.):

Q4 - Please provide the description of your position(s) (e.g. lexicographer, software developer, manager etc.):

Q5 - Do you or your institution or your project use any specialised software to produce dictionary(-like) products?
(A
dictionary writing system - DWS - is a piece of software for writing and
producing a dictionary. It might include an editor, a database, a Web interface
and various management tools - for allocating work etc. Typically, it operates with a
dictionary grammar, which specifies the structure of the dictionary. As an example - IDM DPS, iLex, T-lex, ABBYY Lingvo etc. are well-known
dictionary writing systems, SDL Multiterm would count as a DWS, Microsoft Word would not.)

	No	Yes
Click and write option 1:	○	○

IF (1) Q5a = [2]

Q6 - Please provide the name of the Dictionary Writing System(s) and a URL, bibliographic references or descriptions.
(If there is no URL or bibliographic reference,
please describe the software in one sentence. If you use more than one piece of software please
describe all of them. Bibliographic references can also be to texts in
languages other than English. If you use well-known commercial products – IDM
DPS, iLex, T-lex, ABBYY Lingvo etc. – provide only the name.)

IF (1) Q5a = [2]

Q7 - For which ongoing or finished projects is/was the Dictionary Writing System used? Please provide references for all (or most recent) projects.

(Please provide URLs or bibliographical
references. If these are not available please provide short descriptions of the
projects. Maximum number of characters is 2,500.)

IF (1) Q5a = [2]

Q8 - Is the Dictionary Writing System that you use a commercial off-the-shelf product? (For example: IDM DPS, iLex, T-lex, ABBYY Lingvo etc. - if you use more than one DWS please go through the questionnaire once per each DWS.)

	No	Yes
Click and write option 1:	○	○

IF (3) Q8a = [1]

Q9 - Was the Dictionary Writing System developed within your institution?

	No	Yes
Click and write option 1:	○	○

IF (4) Q9a = [1]

Q10 - Please provide information about the Dictionary Writing System developers and/or maintainers? (Name of the company, individual developers or the people who take care of the accessibility of the software - if you think it is acceptable to provide that information, otherwise please explain briefly why you cannot provide this information.)

IF (5) Q9a = [2] or Q10a'1'

Q11 - Is the Dictionary Writing System used exclusively at your institution / within your project(s) and nowhere else?

	No	Yes
Click and write option 1:	○	○

IF (6) Q11a = [1]

Q12 - Please provide information about other institutions/projects that also use your Dictionary Writing System if possible. Otherwise please explain briefly why you cannot provide this information.

```
┌─────────────────────────────────────────────────────────────────────┐
│                                                                       │
│                                                                       │
│                                                                       │
│                                                                       │
└─────────────────────────────────────────────────────────────────────┘
```

IF (7) Q11a = [2]

Q13 - Is it – to the best of your knowledge – possible to adapt your Dictionary Writing System for other projects in other institutions?

	No	Yes	I don't know
Click and write option 1:	○	○	○

IF (8) Q13a = [1]

Q14 - What are the reasons? (Typical answers: we do not have the possibility to support other users, the software was designed exclusively for our project, the software is language dependent and similar.)

```
┌─────────────────────────────────────────────────────────────────────┐
│                                                                       │
│                                                                       │
│                                                                       │
│                                                                       │
└─────────────────────────────────────────────────────────────────────┘
```

IF (9) Q13a = [2, 3]

Q15 - If possible please provide the information about the people or institutions who can be consulted about the possible adaptations or the use of software by other parties. Otherwise please explain briefly why you cannot provide this information.

```
┌─────────────────────────────────────────────────────────────────────┐
│                                                                       │
│                                                                       │
│                                                                       │
│                                                                       │
└─────────────────────────────────────────────────────────────────────┘
```

IF (11) Q8a = [2]

Q16 - Did you make any adaptations to make it more suitable for your project(s)?

	No	Yes
Click and write option 1:	○	○

IF (12) Q16a = [2]

Q17 - Please describe the adaptations in a few sentences. (If you can't provide an answer please explain why.)

```

```

Q18 - Do you know what is the copyright status of the Dictionary Writing System or who is its owner?

	No	Yes
Click and write option 1:	○	○

IF (15) Q18a = [1]

Q19 - Please provide the contact for the people who can be consulted about the copyright status and/or the availability of the software. (Name - and email, if possible - of the person who is responsible for the software. Can also be an institutional email.)

```

```

IF (16) Q18a = [2]

Q20 - Please provide the information about the copyright status or ownership of the software. (Provide the name of the copyright owner or the licence: GNU licence, Creative Commons, Apache licence etc.)

```

```

Q21 - Does your DWS include control and validation of dictionary structure or »syntax«? (You can use formal mark-up of dictionary structure such as XML schemas, DTDs or similar mechanisms.)

	No	Yes
Click and write option 1:	○	○

Q22 - Does your DWS include consistency checking according to the style guide? (You can use closed sets with only a limited number of values for certain fields, e.g. grammatical codes, labels, different ordering of headwords is available, you can control defining vocabulary and similar.)

	No	Yes

Click and write option 1:	○	○

Q23 - Does your DWS include complex search functionality? (You can use structural hierarchy, Boolean operators, and combinations of these etc. for searching or filtering the contents, for instance - you can search for all nouns in neuter gender ending in -s.)

	No	Yes
Click and write option 1:	○	○

Q24 - Does your DWS use multiple views functionality? (You can view entry data in a structured, tree-diagram view and/or in a WYSIVYG or preview mode.)

	No	Yes
Click and write option 1:	○	○

Q25 - Can you use templates for common dictionary structures or parts of structures? (e.g. You can start a new entry with the basic structure for different parts-of-speech including a headword, division of senses, examples etc.)

	No	Yes
Click and write option 1:	○	○

Q26 - Does your DWS ensure that no more than one person at a time can edit the same entry?

	No	Yes
Click and write option 1:	○	○

Q27 - Are cross references updated automatically?

	No	Yes
Click and write option 1:	○	○

Q28 - Does your DWS include a real-time spellchecker? Real-time means that you can use it while editing the entry.

	No	Yes
Click and write option 1:	○	○

Q29 - Is it possible to include multimedia (images, sounds etc.) in dictionary entries?

	No	Yes
Click and write option 1:	○	○

Q30 - Does your DWS provide any kind of integrated help, for instance context-sensitive help? ('Yes' means that you can click on icons or similar elements in different parts of the interface which provide explanations how that particular part of the software works.)

	No	Yes
Click and write option 1:	○	○

Q31 - Does your DWS have administrative functionality? (You can manage users, assign entries to different users etc.)

	No	Yes
Click and write option 1:	○	○

Q32 - Does your DWS have workflow functionality? (You can control which entries are in different stages of the editorial process before the final publication.)

	No	Yes
Click and write option 1:	○	○

Q33 - Does your DWS have version control? (You can access previous versions of entries stored by the system.)

	No	Yes
Click and write option 1:	○	○

Q34 - Does your DWS have integrated external visualization functionality? (You can publish final entries directly from your DWS to a web page, mobile app or similar.)

	No	Yes
Click and write option 1:	○	○

Q35 - Does your DWS allow import of data? (You can import dictionary data in XML, CSV or other formats.)

	No	Yes
Click and write option 1:	○	○

IF (20) Q35a = [2]

Q36 - Which formats can you import into your DWS?

Q37 - Does your DWS allow export of data? (You can import dictionary data in XML, CSV or other formats.)

	No	Yes
Click and write option 1:	○	○

IF (21) Q37a = [2]

Q38 - Which formats can you export from your DWS? (You can export dictionary data to XML, PDF, InDesign etc.)

Q39 - Does your DWS work off-line? ('Yes' means you do not need access to the internet to edit entries in your DWS.)

	No	Yes
Click and write option 1:	○	○

Q40 - Do you edit dictionary entries in a standard web browser? ('Yes' means you access and edit your dictionary in Firefox, Internet Explorer, Chrome, Safari etc.)

	No	Yes
Click and write option 1:	○	○

Q41 - Does your DWS allows mass data updates? ('Yes' means that for instance one correction to be applied to a whole set of documents in a single operation.)

	No	Yes
Click and write option 1:	○	○

Q42 - Please provide information about the functionalities of your DWS which you particularly like and/or dislike or you think are important to mention.

Q43 - Does your DWS work on a client-server basis?

	No	Yes

Click and write option 1:	◯	◯

IF (23) Q43a = [2]

Q44 - Which database system is used on the server part? (Expected answers: MySQL, ORACLE etc.)

```

```

IF (23) Q43a = [2]

Q45 - Which operating system can be used on the server part? (Expected answers: Linux, Windows Server etc.)

```

```

IF (23) Q43a = [2]

Q46 - Please describe which technology is used for the server part? (Expected answers: Ruby, Python, Java etc.)

```

```

IF (23) Q43a = [2]

Q47 - Please describe which technology is used for the client part? (Expected answers: JavaScript, JSON, XUL etc.)

```

```

IF (24) Q43a = [1]

Q48 - Please describe which technology is your DWS based on?

```

```

Q49 - Does your Dictionary Writing System include any kind of integration with a Corpus Query System?

	No	Yes
Click and write option 1:	○	○

IF (26) Q49a = [2]

Q50 - Does your system integrate a DWS and a CQS in one piece of software?

	No	Yes
Click and write option 1:	○	○

IF (26) Q49a = [2]

Q51 - Please describe briefly the interaction between your DWS and CQS?

```



```

IF (27) Q5a = [1] or Q9a = [1] or Q16a = [1] or Q17a'1'
or Q49a = [1] or Q51a'1'

Q52 - Do you or your institution(s) or your project(s) use any specialised software to query text corpora?
(A corpus query system - CQS - is a piece of software that lets you see concordances for any word, phrase or grammatical construction in a text corpus – an extensive collection of texts, usually annotated with additional information about words, such as their base form or lemma, part-of-speech category and similar.)

	No	Yes
Click and write option 1:	○	○

IF (28) Q52a = [2]

Q53 - Please provide the name of the Corpus Query System and a URL, bibliographic references or descriptions.
(If there is no URL or bibliographic reference,
please describe the software. If you use more than one piece of software please
describe all of them. Bibliographic references can also be to texts in
languages other than English. If you use well-known commercial products – Sketch Engine, CQS part of T-lex etc. – provide only the name.)

```



```

IF (28) Q52a = [2]

Q54 - For which ongoing or finished projects is/was the software used? Please provide references for all (or most recent) projects.

(Please provide URLs or bibliographical
references. If these are not available please provide short descriptions of the
projects in one sentence. If the same answer applies as in the case of your DWS just write "Same as for the DWS".)

```

```

IF (30) Q54a'1'

Q55 - Is the Corpus Query System you use a commercial product? (Examples: Sketch Engine, CQS in T-lex software etc. If you use two or more systems please fill in a separate questionnaire for each of them.)

	No	Yes
Click and write option 1:	○	○

IF (31) Q55a = [2]

Q56 - Did you make any adaptations to make it more suitable for your project(s)?

	No	Yes
Click and write option 1:	○	○

IF (32) Q56a = [2]

Q57 - Please describe the adaptations.

```

```

IF (33) Q55a = [1]

Q58 - Is the software you use an adaptation of a (well-known) free or open source product? (Examples: IMS Open Corpus Workbench (CWB), noSke(tch Engine) etc.)

	No	Yes
Click and write option 1:	○	○

IF (35) Q58a = [1]

Q59 - Was the software developed within your institution?

	No	Yes
Click and write option 1:	○	○

IF (36) Q59a = [1]

Q60 - Please provide some information about the software developer (or the people who take care of the software)? (Name of the company, individual developers or the people who take care of the accessibility of the software - if you think it is acceptable to provide that information, otherwise please explain briefly why you cannot provide this information.)

IF (37) Q59a = [2] or Q60a'1'

Q61 - Is the Corpus Query System used exclusively at your institution / within your project(s) and nowhere else?

	No	Yes
Click and write option 1:	○	○

IF (38) Q61a = [1]

Q62 - Please provide information about other institutions/projects that also use your Corpus Query System if possible. Otherwise please explain briefly why you cannot provide this information.

IF (39) Q61a = [2] or Q62a'1'

Q63 - Is it – to the best of your knowledge – possible to adapt your Corpus Query System for querying other corpora in other institutions?

	No	Yes	I don't know
Click and write option 1:	○	○	○

IF (40) Q63a = [1]

Q64 - What are the reasons? (Typical answers: we do not have the possibility to support other users, the software was designed exclusively for our project, the software is language dependent and similar.)

IF (41) Q63a = [2, 3]

Q65 - If possible please provide the information about the people or institutions who can be consulted about the possible adaptations or the use of software by other parties. Otherwise please explain briefly why you cannot provide this information.

IF (43) Q58a = [2]

Q66 - Was the Corpus Query System adaptation developed within your institution?

	No	Yes
Click and write option 1:	○	○

IF (44) Q66a = [1]

Q671 - Please provide some information about the software developer (or the people who take care of the software)? (Name of the company, individual developers or the people who take care of the accessibility of the software - if you think it is acceptable to provide that information, otherwise please explain briefly why you cannot provide this information.)

IF (45) Q66a = [2]

Q672 - Is the Corpus Query System adaptation used exclusively at your institution / within your project(s) and nowhere else?

	No	Yes
Click and write option 1:	○	○

IF (46) Q672a = [1]

Q673 - Please provide information about other institutions/projects that also use your Corpus Query System adaptation if possible. Otherwise please explain briefly why you cannot provide this information.

IF (47) Q672a = [2]

Q674 - Is it – to the best of your knowledge – possible to adapt your Corpus Query System adaptation for querying other corpora in other institutions?

	No	Yes	I don't know
Click and write option 1:	○	○	○

IF (48) Q672a = [1]

Q675 - What are the reasons? (Typical answers: we do not have the possibility to support other users, the software was designed exclusively for our project, the software is language dependent and similar.)

```

```

IF (49) Q674a = [2, 3]

Q676 - If possible please provide the information about the people or institutions who can be consulted about the possible adaptations or the use of software by other parties. Otherwise please explain briefly why you cannot provide this information.

```

```

IF (51) Q675a'1' or Q676a'1' or Q64a'1' or Q65a'1'

Q680 - Do you know what is the copyright status of the Corpus Query System or who is its owner?

	No	Yes
Click and write option 1:	○	○

IF (52) Q680a = [1]

Q681 - Please provide the contact for the people who can be consulted about the copyright status and/or the availability of the software. (Name - and email, if possible - of the person who is responsible for the software. Can also be an institutional email.)

```

```

IF (53) Q680a = [2]

Q682 - Please provide the information about the copyright status or ownership of the software. (Provide the name of the copyright owner or the licence: GNU licence, Creative Commons, Apache licence etc.)

<div style="border:1px solid #000; height:140px;"></div>

IF (55) Q682a'1' or Q681a'1'

Q691 - Does your CQS index annotated text, so you can search for specific lemmas or parts of speech?

	No	Yes
Click and write option 1:	○	○

Q692 - Does your CQS enable search for regular-expression-like patterns or other kind of complex search? (Example: you can search with operators like .* and combine searches such as words beginning with »anti.*« and ending in ».*ing« filtered by the category »nouns«, sequences of adjectives followed by a noun, and similar.)

	No	Yes
Click and write option 1:	○	○

Q693 - Does your CQS enable identification of collocations? (Example: which is the most common word used before the noun »table«, based on different statistical scores such as log-likelihood, mutual information score etc.)

	No	Yes
Click and write option 1:	○	○

Q694 - Does your CQS allow the use of APIs to extract data from the corpus? (Example: you can write a script in one of the programming languages which extracts a number of concordances or other data – collocations, lists etc. - from the corpus and stores them into a separate file.)

	No	Yes
Click and write option 1:	○	○

Q695 - Does your CQS allows you to show or export lists (on the basis of the search with regular expressions)? (Example: the list of the most frequent morphological tags used in the corpus etc.)

	No	Yes
Click and write option 1:	○	○

Q696 - Does your CQS have some visualization functionality? (Example: graphs indicating the frequency of searched terms in relation to various metadata information – date of publication, genre etc.)

	No	Yes
Click and write option 1:	○	○

Q697 - Does your CQS allow you to create corpora from the web with an automatic procedure?

	No	Yes
Click and write option 1:	○	○

Q698 - Can your CQS process parallel corpora?

	No	Yes
Click and write option 1:	○	○

Q699 - Does your CQS allow import of texts? (You can import dictionary data in XML, TEI or other formats.)

	No	Yes
Click and write option 1:	○	○

IF (57) Q698a = [2]

Q67 - Which formats can you import into your CQS?

Q690 - Does your CQS work off-line? ('Yes' means you do not need access to the internet to search concordances.)

	No	Yes
Click and write option 1:	○	○

Q68 - Please provide information about the functionalities of your CQS that you think are important and were not included in the questionnaire.

IF (59) Q690a = [1, 2] or Q68a'1'

Q700 - Does your CQS work on a client-server basis?

	No	Yes
Click and write option 1:	○	○

IF (60) Q700a = [2]

Q701 - Which database system is used on the server part? (Expected answers: MySQL, ORACLE etc.)

IF (60) Q700a = [2]

Q702 - Which operating system can be used on the server part? (Expected answers: Linux, Windows Server etc.)

IF (60) Q700a = [2]

Q703 - Please describe which technology is used for the server part? (Expected answers: Ruby, Python, Java etc.)

IF (60) Q700a = [2]

Q704 - Please describe which technology is used for the client part? (Expected answers: JavaScript, JSON, XUL etc.)

IF (61) Q700a = [1]

Q705 - Please provide information about the technical characteristics of your CQS that you think are important and were not included in the questionnaire.

Q69 - Do you have any general comments?
(Please use this text box to describe your
dictionary making process or use of corpora without DWSs or CQSs, e.g. if you
use Word, OpenOffice, Emacs or other general word-processing tools. Or if you want to provide information that couldn't be put into previous
text boxes. Or if you want to express your opinion about the survey.)