Glossa – The Corpus Explorer

Lars Nygaard

Version 0.9

Contents

List of Figures

Chapter 1

Introduction

GLOSSA is a web-based user interface for querying linguistic corpora. It is a front-end for the corpus query engine CWB[1].

The development aims have been to create a user interface that is both user friendly and flexible. These two goals are not entirely compatible, however, and the resulting compromise does not allow the user the full range of expression in the CWB search language. Therefore, a separate interface has been created where search expressions can be entered directly, while still enabling GLOSSA's features for browsing and postprocessing of results. This will only be necessary for very complex queries, and most users will not need it.

1.1 System requirements

Most modern web browsers can be used:

- Internet Explorer
- Mozilla, Firefox, Galeon and the rest of the Mozilla family of browsers
- Opera
- Safari

There are currently two exeptions:

- Konqueror
- Internet Explorer for the Macintosh

Konqueror should be supported in future versions; but Internet Explorer for Mac seems to have fallen out of use and will not be supported.

[1] http://cwb.sf.net

Chapter 2

Querying monolingual corpora

In it's simplest incarnation, the GLOSSA interface should look like figure 2.1. For multilingual corpora and corpora with bibliograpic databases, there will be some additional options; these will be explained later.

If you type a word into the box designated «word 1» and press «Search corpus», a new window will appear[1]. In our example, we are querying a corpus of Northern Sami. Searching for the word «gaskavuohta» (meaning «relation»), gives us a *results page* (figure 3.1). The contents of the results page will be explained in section 3.

2.1 Word options

In addition to simply typing a value in the search field, users can restrict the search further by clicking the «options» button, and selecting values from the

[1] In some browsers, you might have to adjust your settings to allow this.

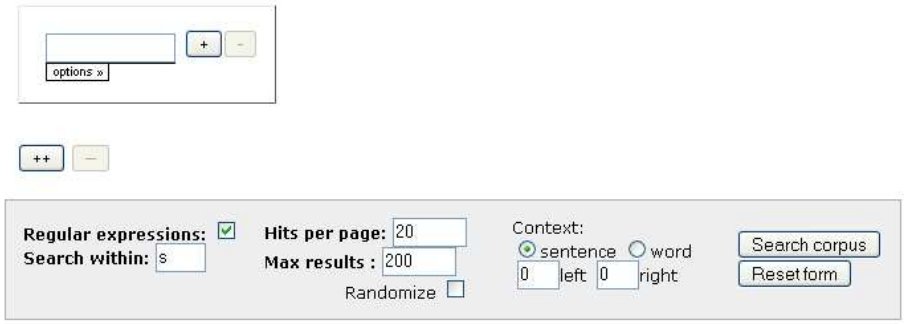


Figure 2.1: Simple interface.

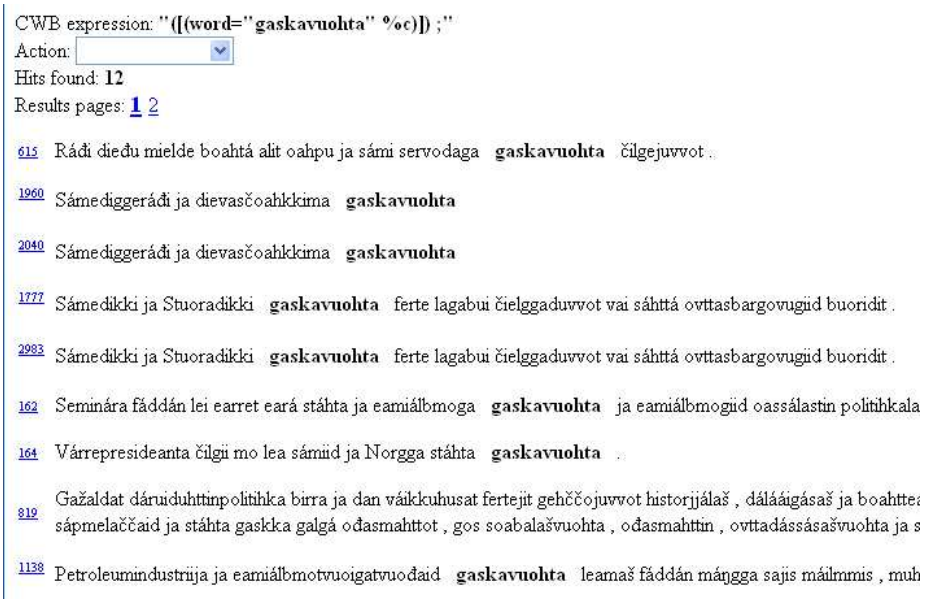


Figure 2.2: Results page.

menu (figure 2.3). When selected, they appear in a box below the options button (figure 2.4). The options in this box can be removed by double-clicking them.

Most options can optionally be negated; in this case they will appear with a prefixed exclamation mark (figures 2.5and 2.6).

Search string options

Start of word if "cat" is entered in the search string box and "start of word" is selected, the program will also return "cats", "category" etc.

End of word if "cat" is entered in the search string box and "end of word" is selected, the program will also return "housecat", "muscat" etc.

Middle of word if "cat" is entered in the search string box and "middle of word" is selected, the program will also return "housecats", "muscatpie" etc.

Case sensitive if "cat" is entered in the search string box and "case sensitive" is selected, the program will return "cat", but not "Cat".

Lemma if "cat" is entered in the search string box and "lemma" is selected, the program will return all forms of the word, i.e. "cat", and "cats".[2]

Exclude the program will not return words that match.

[2]If lemma annotation is present in the corpus.

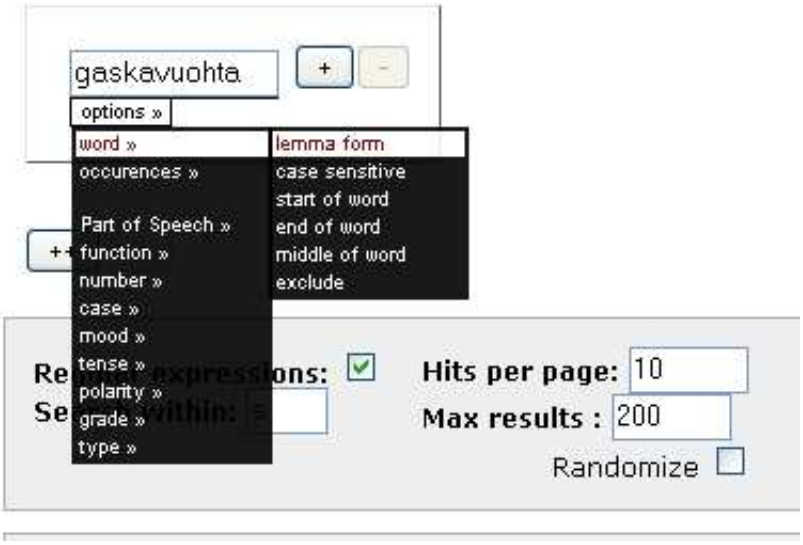


Figure 2.3: Selecting an option.

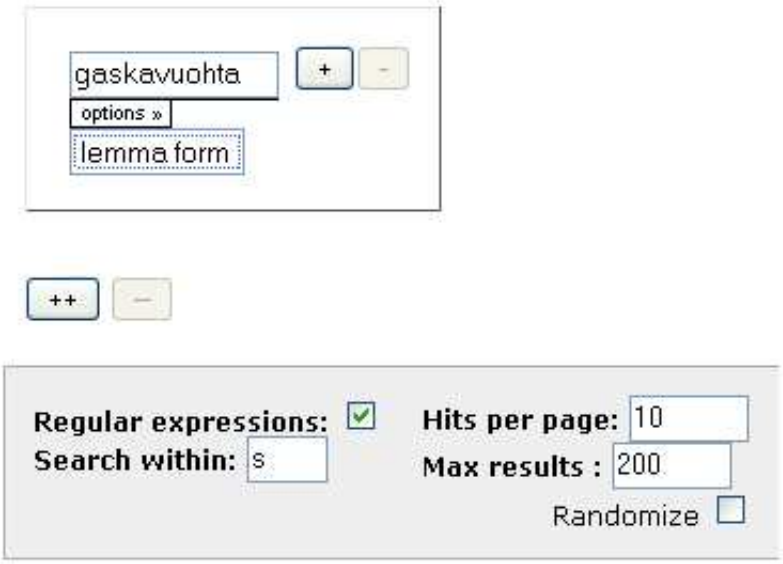


Figure 2.4: Selected option.

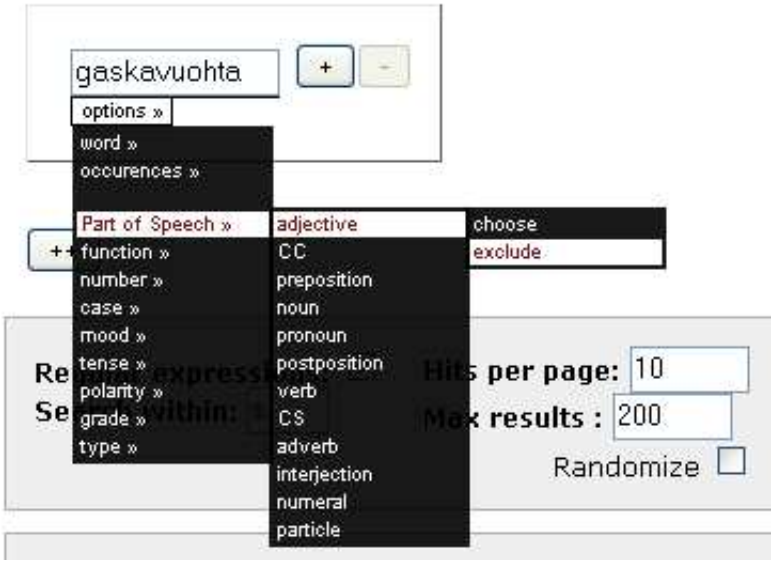


Figure 2.5: Selecting a negated option.

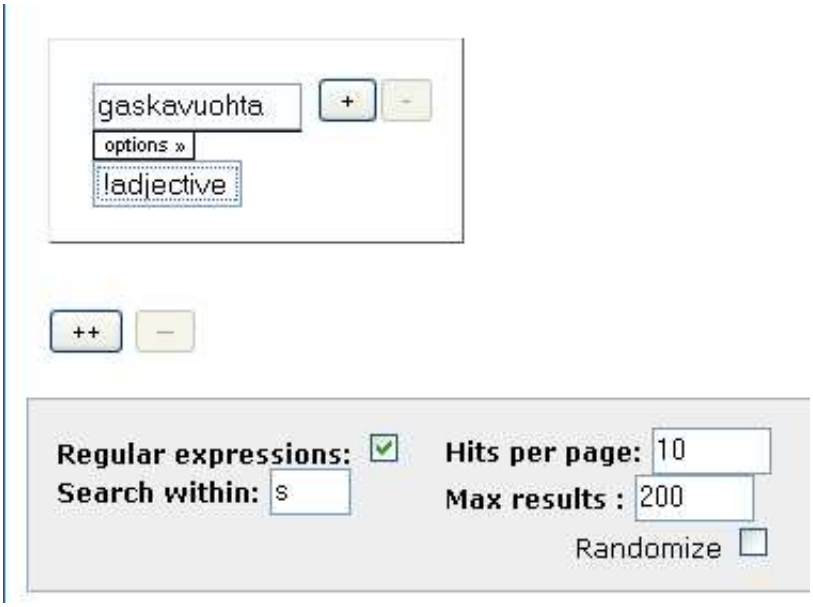


Figure 2.6: A negated option.

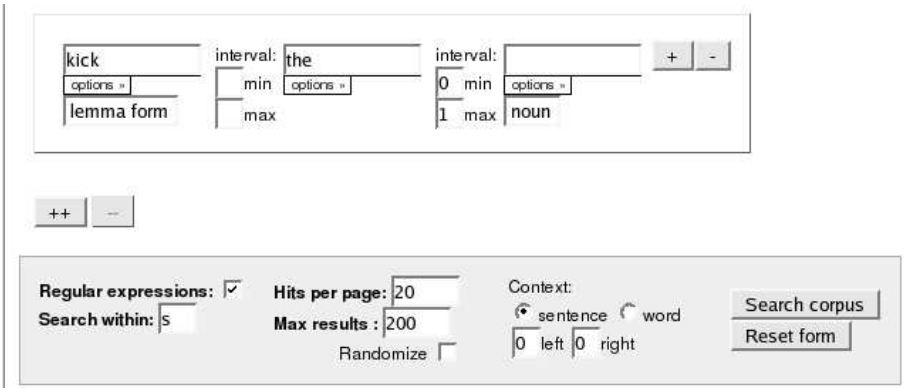


Figure 2.7: Phrase search.

Annotation options

Depending on the degree of linguistic annotation of the corpus you are querying, a range of annotation options are available. Typically, available options will be part-of-speech, morphological features and syntactic functions.

Positive annotation options will be connected by disjunction, negative annotation options with discjunction. Thus selecting "noun", "verb", "!adjective", "!adverb" will return words that are either nouns or verbs, but neither adjectives nor adverbs. More formally:

(2.1) (noun OR verb) AND NOT (adjective OR adverb)

Other options

Occurrences allows specification of how many times the token can occur. If you enter the word «much», and select «one or more», it will match cases like

(2.2) It was much much too cold.

2.2 Intervals and phrases

An important feature of the Glossa system is the ability search very complex phrases. This is done by adding more word boxes, and optinally specifying the lengths of the interwals between them. More word boxes can be added or removed by clicking the buttons with plus or minus signs, on the right side of the screen.

The minimum and maximum interval specifies the number of unspecified words between two query words. If both are left empty, it is assumed that no unspecified tokens can come between the query tokens (i.e. max: zero, min: zero). If the minimum interval is specified, but not the maximum, unlimited

maximum interval is assumed. Conversely, if the maximum interval is specified, but not the minimum, a minimum interval of zero is assumed.

Figure 2.7 demonstrates a search for the following phrase:

(2.3) the lemma "kick"

followed directly by the word "the"

an interval of zero or one unspecified tokens

followed by a noun

The following sentences are some of the ones that matched in a corpus of English texts:

(2.4) He even repeatedly **kicked the piano**, which he used to be so careful of.

(2.5) As I had walked part of the way through the fields, I **kicked the caked mud** off my shoes before going in.

(2.6) People who'd never smoked in their lives hadn't the faintest idea of how difficult it was trying to **kick the habit**, he thought morosely.

(2.7) They wrenched my arms across my back, forced me down into the cranberry heather and punched and **kicked the way** they were trained.

(2.8) Stop to **kick the snow** off my boots.

(2.9) The foreigners whispered to each other, **kicked the blubber** and felt the skins and the walrus tusks.

2.2.1 Additional phrases

In some cases, it can be useful to join two different queries in the same result set. This can be done by adding additional phrases, with the buttion marked «++» (these can be removed by clicking the button marked «- -»).

For example, figure 2.8 demonstrates a search for both «coin museum» and «museum for coins».

2.3 General options

Below the actual word and phrase queries, we find options for the entire search.

Using regular expression

If the regular expressions box is checked, user input will be interpreted as regular expressions: i.e. "." will be interpreted as "one arbitrary character". If it is unchecked, all regular expression characters will be *escaped*: i.e. "." will be interpreted as a period. The regular expression vocabulary is described in Section B.

Queries can sometimes be created faster by typing regular expressions than by selecting items from the menu (eg. typing "house.*" instead of typing "house" and then selecting "start of word" from the menu; typing "house|building" instead of using two query rows). Also, more complex regular expressions cannot be created from the menu.

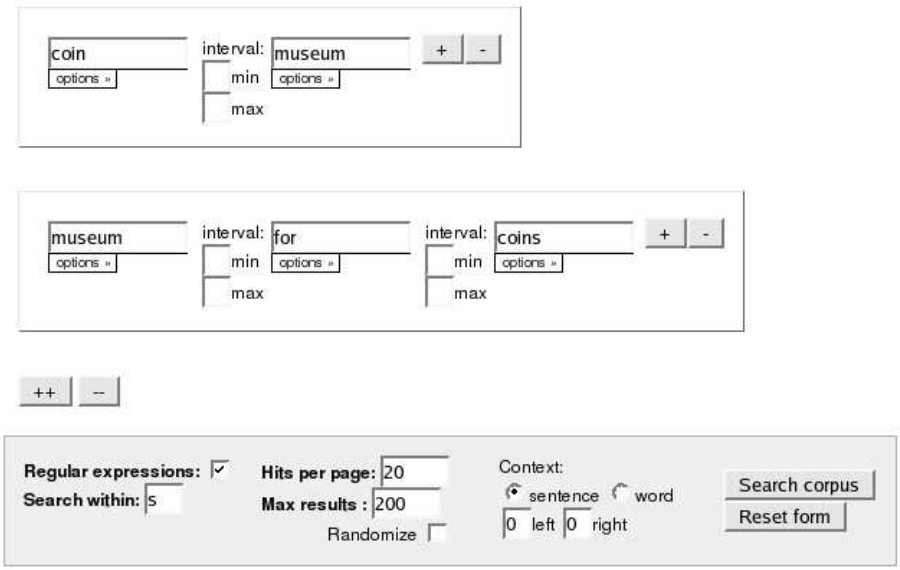


Figure 2.8: Several phrases.

Search within

This parameter restricts the matching of searches containing arbitrary tokens. It can be set to:

's' where all matches will be within the same s-unit. If you corpus contains additional structural markup, like paragraphs, you can use that as well. Refer to the documentation or administrator for the corpus.

integer where all matches will be within the specified number of tokens

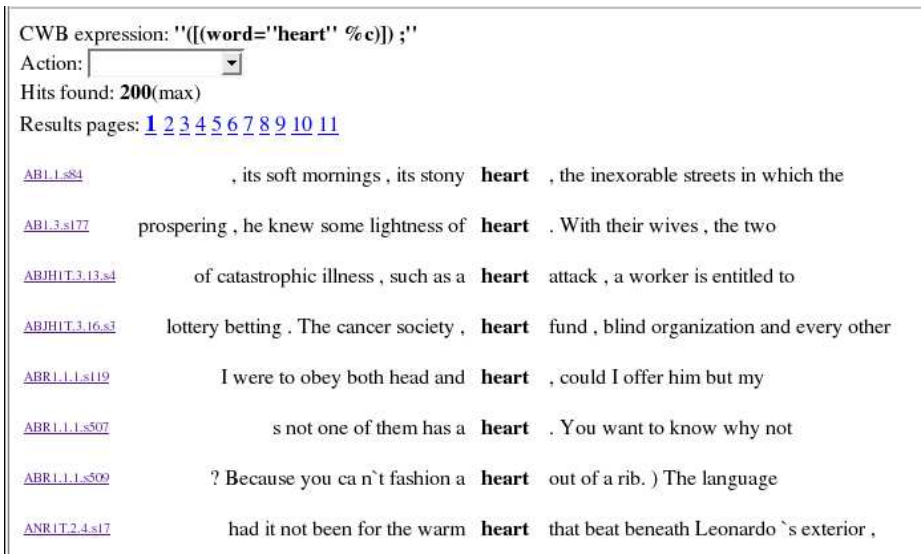
If you set the «search within» parameter to 's' and search for «kick» and «bucket» with an unlimited number of unspecified tokens in the phrase query, you will only get results where «kick» and «bucket» is in the same paragraph.

Results per page

The search results are divided into a number of pages; the number of results on each page can be adjusted here.

Number of results

Searching for common words in large corpora can be slow. Restricting the total number of results can improve response times. Unless 'randomize' is selected, the first hits in the corpus will be displayed.


CWB expression: "([(word="heart" %c)]) ;"
Action:
Hits found: 200(max)
Results pages: 1 2 3 4 5 6 7 8 9 10 11

AB1.1.s84	, its soft mornings , its stony	heart	, the inexorable streets in which the
AB1.3.s177	prospering , he knew some lightness of	heart	. With their wives , the two
ABJH1T.3.13.s4	of catastrophic illness , such as a	heart	attack , a worker is entitled to
ABJH1T.3.16.s3	lottery betting . The cancer society ,	heart	fund , blind organization and every other
ABR1.1.1.s119	I were to obey both head and	heart	, could I offer him but my
ABR1.1.1.s507	s not one of them has a	heart	. You want to know why not
ABR1.1.1.s509	? Because you ca n`t fashion a	heart	out of a rib.) The language
ANR1T.2.4.s17	had it not been for the warm	heart	that beat beneath Leonardo `s exterior ,

Figure 2.9: A KWIC concordance.

Context

There are two ways of specifying contexts size shown in the search results: by number of sentences or by number of tokens. If we select `s-units`, the `left` and `right` boxes specify the number of sentences to the left and right of the matching sentence.

Similarly, if we select tokens, we specify the number of tokens to the left and right to the matching phrase. Also, the results are displayed as a traditional KWIC concordance. Figure 2.9 contains a results page with seven words on both left and right side.

2.4 Meta information

If your corpus contains meta information (predefined subcorpora, bibliograpic information etc.), GLOSSA supports both viewing this information (described in 3) and restricting searches according to it.

If available, meta-data restrictions will appear below the general options, as illustrated in figure 2.10. Restriction classes can be hidden[3] or displayed with the «+» and «-» buttons.

There are three types of restriction classes: tables, ranges and checkboxes. Using ranges («publication date» in 2.10) and checkboxes («translation» in 2.10) should be straighforward. The tables are used as follows:

- items are moved from one column to the other eiter by double-clicking

[3] Hiding a restriction class will also reset it, i.e. no constraint on the search will be created.

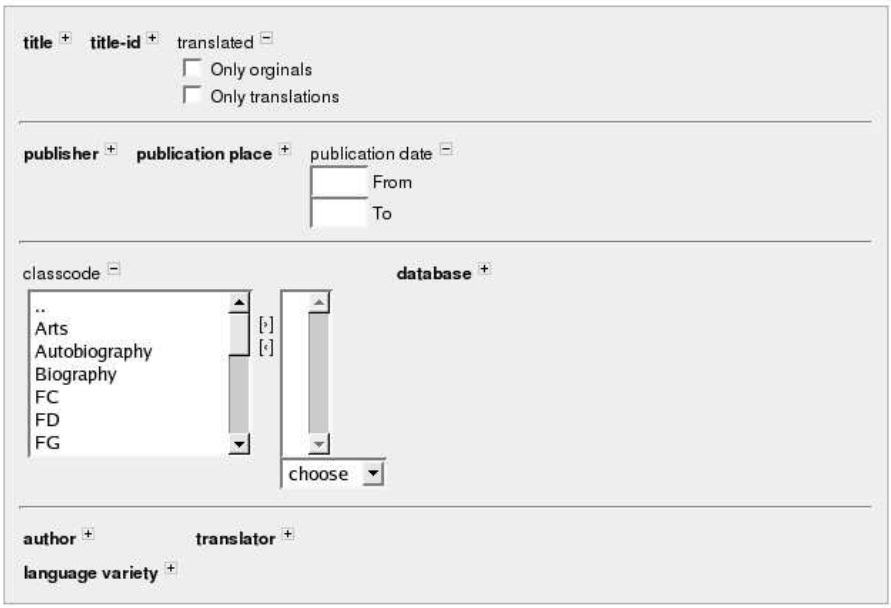


Figure 2.10: Meta-data restrictions.

them, or by marking one or more and clicking the appropriate arrow button.

- the items in the right column are selected, and the menu under that column is used to specify whether those items should be excluded or chosen. If they are chosen, *only* matches from texts with one of those attributes will be returned; if they are excluded, those matches will *not* be returned.

Figure 2.11 will return matches from texts that:

- has the classcode «Arts» or «Autobiography»
- is *not* published in «Amsterdam» or «Hildesheim»
- is published between 1970 and 1980
- is an original (i.e. not a translation)

To the right of the screen the user will find controls to:

Show texts i.e. display a list of texts that will be searched with the current meta-data configuration.

Save subcorpus i.e. save the current meta-data configuration for later use.

Choose subcorpus i.e. select a previously saved configuration.

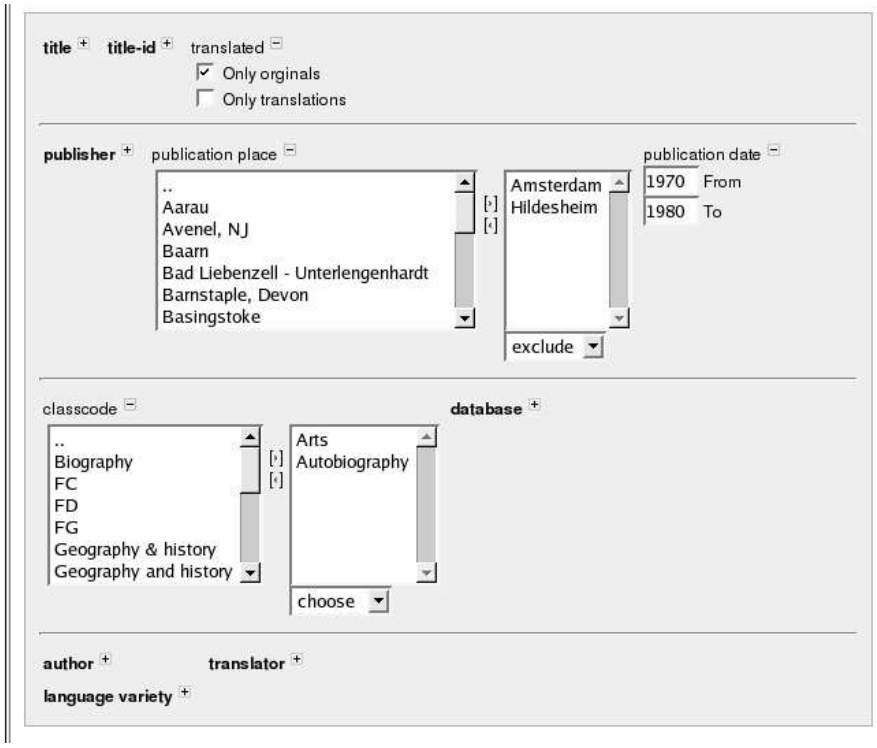


Figure 2.11: Restrictions on meta-data.

Chapter 3

Browsing results

3.1 The results pages

The results page consists of:

- The CWB search string
- A list of available actions for further processing of the result set (described in section 3.2).
- The number of matches returned.
- A list of results pages, with the current page shown in bold.
- The results themselves. Each result then consists of:
  - The sentence id. If this id is clicked, a window appears showing meta-information about the text in which the sentence appears. Additionally, it shows more context (and the user can set the context size to an arbitrary large number).
  - Left context
  - The matching phrase (in bold)
  - Right context
  - Linguistic annotation of each word in the result set; displayed when the mouse is moved over the word

3.2 Processing results

In this section, actions for processing results are presented. One action is only applicable to multilingual corpora – co-occurence statistics – and is presented in section 4.3.

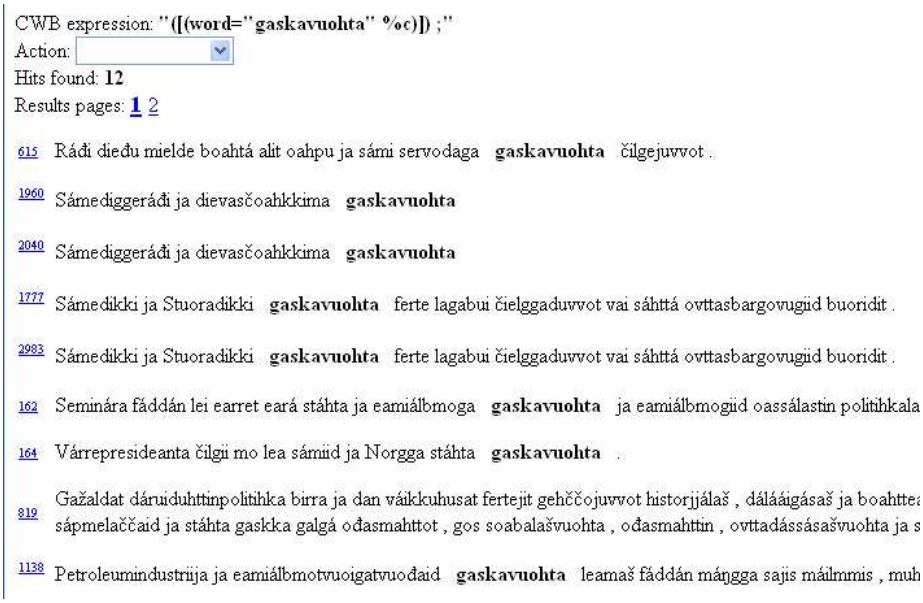


Figure 3.1: Results page.

3.2.1 Sort

The sorting function applies to the order of the matches in the results set. The set can be randomized, or sorted alphabetically, according to the source corpus hits, by

- left context
- right context
- matching phrase
- sentence id

When sorting by context or matching phrase, the sorting can be done according to:

- word form
- lemma
- part-of-speech
- any combination of the above

By default, context sorting is done according to the token that is closest to the matching phrase, but the position in context can be set higher by the user.

If the search criteria of two hits are identical, the secondary search criterion applies, with the same options as the primary criterion.

The setup in figure 3.2 will sort the results

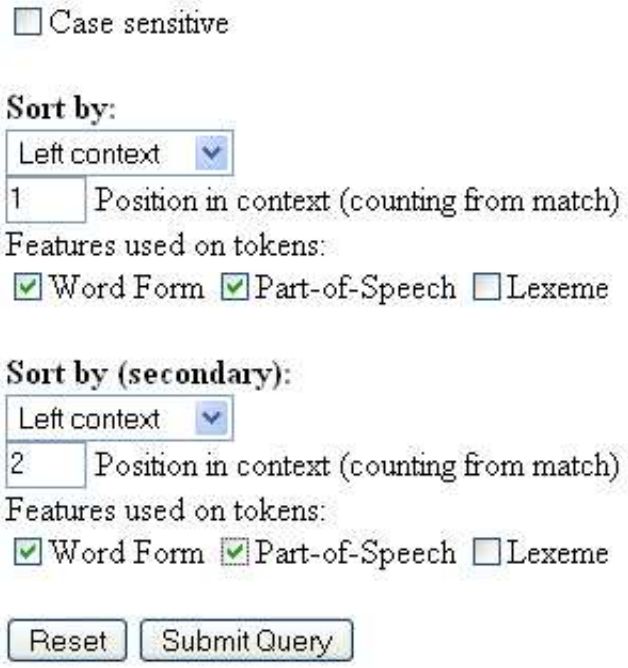


Figure 3.2: Sorting options.

1. by the word form on the first word to the left of the match
2. by the part-of-speech of the first word to the left
3. by the word form of the second word to the left
4. by the part-of-speech of the second word to the left

If all theese criteria are equal, the results will appear in the original order. The start of a result set sorted according to those options are shown in figure 3.3.

3.2.2 Lexical statistics

Statistics can be compiled for:

- word form
- lemma
- part-of-speech
- any combination of the above

The results can be presented in any of the following data formats:

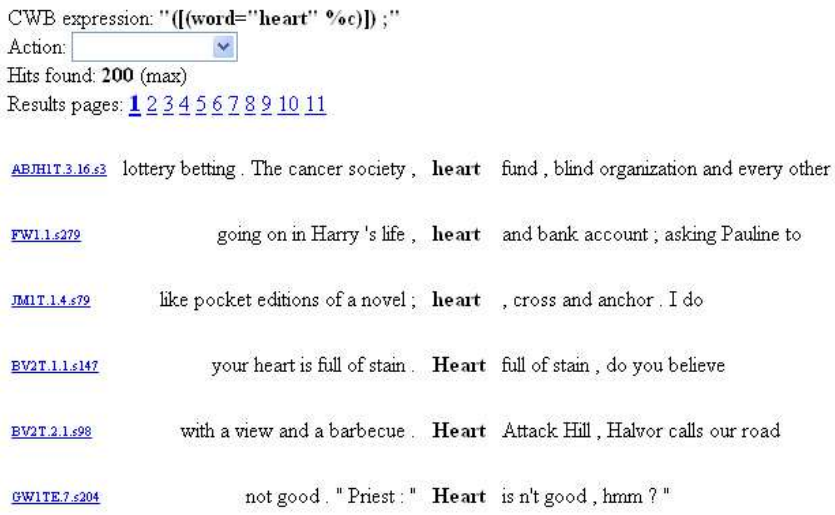


Figure 3.3: Sorted result set.

- HTML
- Tab-separated values
- Comma-separated values
- Excel spreadsheet
- Histogram
- Pie chart

Count

This action generates statistics over the matching phrase in the source corpus.
Statistics can be generated for

- word form
- lemma
- part-of-speech
- any combination of the above

collocations

The collocation function compiles statistics of tokens occurring within a user specified context window of the matching phrase.

The available collocation statistics for bigrams are:

- Frequency (no association measure)
- Dice coefficient
- Fisher's exact test
- Log-likelihood ratio
- Mutual information
- Pointwise mutual information
- Odds ratio
- Phi coefficient
- T-score
- Pearson's chi squared test

The available collocation statistics for trigrams are:

- Frequency (no association measure)
- Log-likelihood ratio

The association measures are described in the Ngram Statistics Package documentation `http://search.cpan.org/dist/Text-NSP/Docs/Measures.pod`.

Note that only first word in the matching phrase is used. Thus if any of the matching phrases contain more than one word, the right-side statistics will contain errors.

3.2.3 Deleting hits

3.2.4 Saving result sets

The entire result set can be saved. This can be done in either of two ways:

- download to disk
- store on server

Download

Optionally, additional meta-data may be included in the downloaded result set. The text in the result set can include

- word form
- lemma
- part-of-speech
- any combination of the above

Store on server

3.2.5 Meta-data

List texts

Information about the text that where applicable for matches in the query can be displayed using this action. Note that this is the texts for which matches could have been returned; not the ones from which matches were actually returned.

Distribution

This function displays the number of hits, sorted according to the categories of meta-data available, including information about the number of hits per thousand words.

Chapter 4

Querying multilingual corpora

Exploring multilingual corpora with GLOSSA is very similar to exploring monolingual ones:

- some extra options in the search builder
- some extra information in the results page
- one extra option for processing results

4.1 Phrase options

If the corpus is multilingual, each phrase in the search builder has some extra options (figure 4.1).

Notice first that each phrase has an option for *language*. The language selected for the first phrase always constitutes the BASE CORPUS. This corpus will be searched first, and search expressions for all other languages – considered ALIGNED CORPORA – will be mached against the aligned regions of the maches of the base corpus. When you change the language of the first phrase, you change which corpus is considered base corpus.

Optional alignment

Glossa will display aligned regions for all aligned corpora. If you leave the search options empty, it will by default create a constraint of *at least one, unspecified token*. This means that if there are parts of the base corpus that does not have aligned regions in one of the aligned corpora, results from those regions will not return matches. This behaviour can be changed by clicking on the «optional alignment» checkbox: The program will return matches from the base corpus, even if there is no aligned regions for that aligned corpus.

Negation

Search phrases for aligned corpora can be specified to be negative or positive – using the menu next to the language selector – while base corpus phrases can only be positive.[1]

[1] This is a result of limitations in CQP.

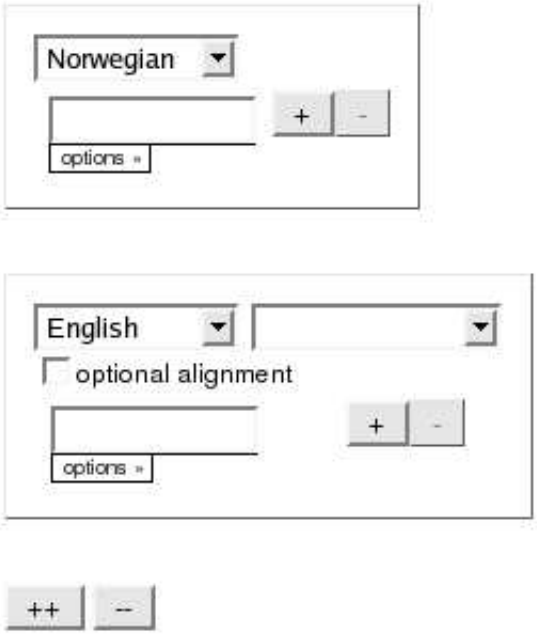


Figure 4.1: Multilingual search.

Relation between query phrases in aligned corpora

If more than one query phrases (see Section **??**) are specified for the same aligned corpus, a new menu appears, allowing us to set the logical relation between them:

conjunction all of the phrases must occur

disjunction at least one phrases must occur

You can select more than one phrases for the base corpus as well, but these can only be disjuncted.[2]

4.2 Browsing results

The results pages are identical to the results pages for monolingual corpora, except that aligned regions appear under each base corpus match, in gray color (figure 4.2).

[2]This is a result of limitations in CQP.

Figure 4.2: Results, multilingual corpus.

4.3 Processing results

The co-occurrence functions provide statistics of the words in the target corpus hits.

4.4 Miscellanea

When searching for common words, it is recommended to use 'randomize', since it will generally be faster.

Context size for aligned corpora cannot be set; it is always the region or regions aligned to the matching sentence.

Meta-information applies to the base corpus, not the aligned corpora.

Chapter 5

Querying speech corpora

Chapter 6

Creating frequency tables

The normal search interface can be used to create many kinds of frequency tables, particularly if one takes care to set the 'max results' value appropriately. However, for efficiency reasons, very general frequency lists cannot be created. Therefore, there is a separate interface for creating tables of word frequencies from entire corpora (figure 6.1).

The user can specify what to include in the table (this should be familiar from section 3): word form, lemma form, part-of-speech label, or a combination of theese.[1]

There are several kinds of restrictions that the user can put on the compilation:

- POS filter: Only include words of a particular part-of-speech
- Cutoff: Only include this number of words (i.e. the 1000 most frequent words)
- Meta-data: A subcorpus can be created according to meta-data restrictions. This is done in the same way as described in section 2.4.

If there are meta-data restrictions, the compilation will include a *contrastive frequency* count; the table will also, by default, be sorted by decreasing frequency contrast. This means, essentially, that the expected frequency of each word (computed from the rest of the corpus) is contrasted with the actual frequency in the selected subcorpus.

In the example in figure 6.2, Norwegian texts with the topic «health» has been selected. The first word, «pasient» ('patient') occurs 1237 times in those texts; it occurs 1661 times in the entire corpus, and thus we would expect it to occur 7664 times (this depends on the total number of tokens in the corpus). Thus the frequency contrast is 6003. Other words with high frequency contrast is «legemiddel» ('medication'), «sykehus» ('hospital'), «apotek» ('pharmacy') etc.

Note that all frequency compilations can take up to several minutes, depending on the size of the corpus.

[1] These choises are ignored when meta-data restrictions are selected: That will always give lemma and part-of-speech label.

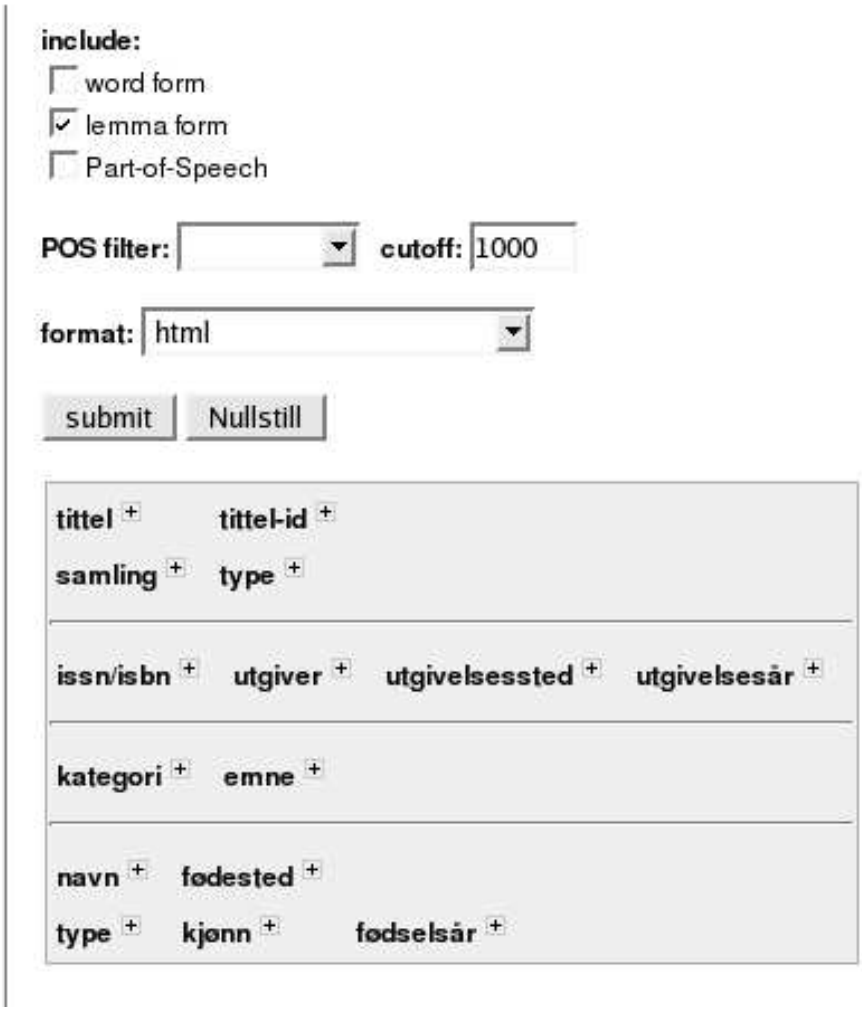


Figure 6.1: Interface for frequency tables.

#	word	freq	glob freq	expected	diff
1	pasient__s	1237	1661	7664	6003
2	legemiddel__s	869	43	5384	5341
3	sykehus__s	720	1272	4461	3189
4	apotek__s	503	13	3116	3103
5	behandling__s	836	2333	5179	2846
6	tuberkulose__s	466	119	2887	2768
7	lege__s	646	2118	4002	1884
8	regionsykehus__s	289	4	1790	1786
9	kreft__s	314	293	1945	1652
10	fylkeskommune__s	237	118	1468	1350
11	medisinsk__adj	290	735	1796	1061
12	grossist__s	153	19	947	928
13	preparat__s	150	31	929	898
14	helsetjeneste__s	156	128	966	838
15	tuberkulosekontroll__s	136	7	842	835

Figure 6.2: Frequency contrasts for health-related Norwegian texts.

Chapter 7

Querying treebanks

Appendix A

Acknowledgements

The development has been done at The Text Laboratory, University of Oslo; partially financed by the SPRIK project `http://www.hf.uio.no/forskningsprosjekter/sprik/`.

Several people have given valuable input to the development: Janne Bondi Johannessen, Kristin Hagen, Åshild Søfteland, Cathrine Fabricius Hansen, Stig Johansson, Hilde Hasselgård, Hans Petter Helland, Anders Nøklestad, Elisabeth Lien, Ruth Vatvedt Fjeld, Bergljot Behrens, Hilde C. F. Haug, Ingunn Indrebø Ims, Signe Laake, Inger Margrethe Seim, Torgrim Solstad, Wiebke Ramm.

Other corpus interfaces that has provided inspiration are:

- CorpusEye, developed at The University of Southern Denmark (http://corp.hum.sdu.dk `http://corp.hum.sdu.dk`); designed by Eckhard Bick and by Poul Henriksen and Nikolaj Hald Nielsen.
- BNCweb, developed at Zurich University (http://homepage.mac.com/bncweb `http://homepage.mac.com/bncweb`), by Hans-Martin Lehmann, Sebastian Hoffmann and Peter Schneider
- Interfaces by Paul Meurer and Knut Hofland, at the Aksis Center, University of Bergen

Appendix B

Regular expressions

A full account of the regular expressions used by CWB can be found on the IMS website `http://www.ims.uni-stuttgart.de/projekte/CorpusWorkbench/CQPUserManual/HTML/`.

B.1 Optionality

The period (".") represents any character. Thus `.ats` will match "cats", "mats", "bats" etc.

A list of alternative characters can be represented with square brackets: `[cm]ats` will match either "cats" or "mats".

A list of alternative strings can be represented with the vertical bar: `cats|mats` will again match either "cats" or "mats".

B.2 Occurrences

The number of times characters can occur can be specified with the following operators:

? `cats?` matches both "cat" and "cats"

***** `the*` matches "th", "the", "thee" etc.

+ `the+` matches "the", "thee", etc.

{n,n} `the{1,2}` matches "the" or "thee".

B.3 Escaping operators

All the regular expression operators can be searched for; they are interpreted literally if they are prefixed by a backslash. Thus `\.` maches a period in the corpus, and `\?` matches a question mark.

Appendix C

CE licencing

GNU GENERAL PUBLIC LICENSE

Version 2, June 1991

Preamble

The licenses for most software are designed to take away your freedom to share and change it. By contrast, the GNU General Public License is intended to guarantee your freedom to share and change free software–to make sure the software is free for all its users. This General Public License applies to most of the Free Software Foundation's software and to any other program whose authors commit to using it. (Some other Free Software Foundation software is covered by the GNU Lesser General Public License instead.) You can apply it to your programs, too.

When we speak of free software, we are referring to freedom, not price. Our General Public Licenses are designed to make sure that you have the freedom to distribute copies of free software (and charge for this service if you wish), that you receive source code or can get it if you want it, that you can change the software or use pieces of it in new free programs; and that you know you can do these things.

To protect your rights, we need to make restrictions that forbid anyone to deny you these rights or to ask you to surrender the rights. These restrictions translate to certain responsibilities for you if you distribute copies of the software, or if you modify it.

For example, if you distribute copies of such a program, whether gratis or for a fee, you must give the recipients all the rights that you have. You must make sure that they, too, receive or can get the source code. And you must show them these terms so they know their rights.

We protect your rights with two steps: (1) copyright the software, and (2) offer you this license which gives you legal permission to copy, distribute and/or modify the software.

Also, for each author's protection and ours, we want to make certain that everyone understands that there is no warranty for this free software. If the software is modified by someone else and passed on, we want its recipients to know that what they have is not the original, so that any problems introduced by others will not reflect on the original authors' reputations.

Finally, any free program is threatened constantly by software patents. We wish to avoid the danger that redistributors of a free program will individually obtain patent licenses, in effect making the program proprietary. To prevent this, we have made it clear that any patent must be licensed for everyone's free use or not licensed at all.

The precise terms and conditions for copying, distribution and modification follow.

TERMS AND CONDITIONS FOR COPYING, DISTRIBUTION AND MODIFICATION

0. This License applies to any program or other work which contains a notice placed by the copyright holder saying it may be distributed under the terms of this General Public License. The "Program", below, refers to any such program or work, and a "work based on the Program" means either the Program or any derivative work under copyright law: that is to say, a work containing the Program or a portion of it, either verbatim or with modifications and/or translated into another language. (Hereinafter, translation is included without limitation in the term "modification".) Each licensee is addressed as "you".

Activities other than copying, distribution and modification are not covered by this License; they are outside its scope. The act of running the Program is not restricted, and the output from the Program is covered only if its contents constitute a work based on the Program (independent of having been made by running the Program). Whether that is true depends on what the Program does.

1. You may copy and distribute verbatim copies of the Program's source code as you receive it, in any medium, provided that you conspicuously and appropriately publish on each copy an appropriate copyright notice and disclaimer of warranty; keep intact all the notices that refer to this License and to the absence of any warranty; and give any other recipients of the Program a copy of this License along with the Program.

You may charge a fee for the physical act of transferring a copy, and you may at your option offer warranty protection in exchange for a fee.

2. You may modify your copy or copies of the Program or any portion of it, thus forming a work based on the Program, and copy and distribute such modifications or work under the terms of Section 1 above, provided that you also meet all of these conditions:

a) You must cause the modified files to carry prominent notices stating that you changed the files and the date of any change.

b) You must cause any work that you distribute or publish, that in whole or in part contains or is derived from the Program or any part thereof, to be licensed as a whole at no charge to all third parties under the terms of this License.

c) If the modified program normally reads commands interactively when run, you must cause it, when started running for such interactive use in the most

ordinary way, to print or display an announcement including an appropriate copyright notice and a notice that there is no warranty (or else, saying that you provide a warranty) and that users may redistribute the program under these conditions, and telling the user how to view a copy of this License. (Exception: if the Program itself is interactive but does not normally print such an announcement, your work based on the Program is not required to print an announcement.)

These requirements apply to the modified work as a whole. If identifiable sections of that work are not derived from the Program, and can be reasonably considered independent and separate works in themselves, then this License, and its terms, do not apply to those sections when you distribute them as separate works. But when you distribute the same sections as part of a whole which is a work based on the Program, the distribution of the whole must be on the terms of this License, whose permissions for other licensees extend to the entire whole, and thus to each and every part regardless of who wrote it.

Thus, it is not the intent of this section to claim rights or contest your rights to work written entirely by you; rather, the intent is to exercise the right to control the distribution of derivative or collective works based on the Program.

In addition, mere aggregation of another work not based on the Program with the Program (or with a work based on the Program) on a volume of a storage or distribution medium does not bring the other work under the scope of this License.

3. You may copy and distribute the Program (or a work based on it, under Section 2) in object code or executable form under the terms of Sections 1 and 2 above provided that you also do one of the following:

a) Accompany it with the complete corresponding machine-readable source code, which must be distributed under the terms of Sections 1 and 2 above on a medium customarily used for software interchange; or,

b) Accompany it with a written offer, valid for at least three years, to give any third party, for a charge no more than your cost of physically performing source distribution, a complete machine-readable copy of the corresponding source code, to be distributed under the terms of Sections 1 and 2 above on a medium customarily used for software interchange; or,

c) Accompany it with the information you received as to the offer to distribute corresponding source code. (This alternative is allowed only for noncommercial distribution and only if you received the program in object code or executable form with such an offer, in accord with Subsection b above.)

The source code for a work means the preferred form of the work for making modifications to it. For an executable work, complete source code means all the source code for all modules it contains, plus any associated interface definition files, plus the scripts used to control compilation and installation of the executable. However, as a special exception, the source code distributed need not include anything that is normally distributed (in either source or binary form) with the major components (compiler, kernel, and so on) of the operating system on which the executable runs, unless that component itself accompanies the executable.

If distribution of executable or object code is made by offering access to copy from a designated place, then offering equivalent access to copy the source code from the same place counts as distribution of the source code, even though third parties are not compelled to copy the source along with the object code.

4. You may not copy, modify, sublicense, or distribute the Program except as expressly provided under this License. Any attempt otherwise to copy, modify, sublicense or distribute the Program is void, and will automatically terminate your rights under this License. However, parties who have received copies, or rights, from you under this License will not have their licenses terminated so long as such parties remain in full compliance.

5. You are not required to accept this License, since you have not signed it. However, nothing else grants you permission to modify or distribute the Program or its derivative works. These actions are prohibited by law if you do not accept this License. Therefore, by modifying or distributing the Program (or any work based on the Program), you indicate your acceptance of this License to do so, and all its terms and conditions for copying, distributing or modifying the Program or works based on it.

6. Each time you redistribute the Program (or any work based on the Program), the recipient automatically receives a license from the original licensor to copy, distribute or modify the Program subject to these terms and conditions. You may not impose any further restrictions on the recipients' exercise of the rights granted herein. You are not responsible for enforcing compliance by third parties to this License.

7. If, as a consequence of a court judgment or allegation of patent infringement or for any other reason (not limited to patent issues), conditions are imposed on you (whether by court order, agreement or otherwise) that contradict the conditions of this License, they do not excuse you from the conditions of this License. If you cannot distribute so as to satisfy simultaneously your obligations under this License and any other pertinent obligations, then as a consequence you may not distribute the Program at all. For example, if a patent license would not permit royalty-free redistribution of the Program by all those who receive copies directly or indirectly through you, then the only way you could satisfy both it and this License would be to refrain entirely from distribution of the Program.

If any portion of this section is held invalid or unenforceable under any particular circumstance, the balance of the section is intended to apply and the section as a whole is intended to apply in other circumstances.

It is not the purpose of this section to induce you to infringe any patents or other property right claims or to contest validity of any such claims; this section has the sole purpose of protecting the integrity of the free software distribution system, which is implemented by public license practices. Many people have made generous contributions to the wide range of software distributed through that system in reliance on consistent application of that system; it is up to the author/donor to decide if he or she is willing to distribute software through any other system and a licensee cannot impose that choice.

This section is intended to make thoroughly clear what is believed to be a consequence of the rest of this License.

8. If the distribution and/or use of the Program is restricted in certain countries either by patents or by copyrighted interfaces, the original copyright holder who places the Program under this License may add an explicit geographical distribution limitation excluding those countries, so that distribution is permitted only in or among countries not thus excluded. In such case, this License incorporates the limitation as if written in the body of this License.

9. The Free Software Foundation may publish revised and/or new versions

of the General Public License from time to time. Such new versions will be similar in spirit to the present version, but may differ in detail to address new problems or concerns.

Each version is given a distinguishing version number. If the Program specifies a version number of this License which applies to it and "any later version", you have the option of following the terms and conditions either of that version or of any later version published by the Free Software Foundation. If the Program does not specify a version number of this License, you may choose any version ever published by the Free Software Foundation.

10. If you wish to incorporate parts of the Program into other free programs whose distribution conditions are different, write to the author to ask for permission. For software which is copyrighted by the Free Software Foundation, write to the Free Software Foundation; we sometimes make exceptions for this. Our decision will be guided by the two goals of preserving the free status of all derivatives of our free software and of promoting the sharing and reuse of software generally.

NO WARRANTY

11. BECAUSE THE PROGRAM IS LICENSED FREE OF CHARGE, THERE IS NO WARRANTY FOR THE PROGRAM, TO THE EXTENT PERMITTED BY APPLICABLE LAW. EXCEPT WHEN OTHERWISE STATED IN WRITING THE COPYRIGHT HOLDERS AND/OR OTHER PARTIES PROVIDE THE PROGRAM "AS IS" WITHOUT WARRANTY OF ANY KIND, EITHER EXPRESSED OR IMPLIED, INCLUDING, BUT NOT LIMITED TO, THE IMPLIED WARRANTIES OF MERCHANTABILITY AND FITNESS FOR A PARTICULAR PURPOSE. THE ENTIRE RISK AS TO THE QUALITY AND PERFORMANCE OF THE PROGRAM IS WITH YOU. SHOULD THE PROGRAM PROVE DEFECTIVE, YOU ASSUME THE COST OF ALL NECESSARY SERVICING, REPAIR OR CORRECTION.

12. IN NO EVENT UNLESS REQUIRED BY APPLICABLE LAW OR AGREED TO IN WRITING WILL ANY COPYRIGHT HOLDER, OR ANY OTHER PARTY WHO MAY MODIFY AND/OR REDISTRIBUTE THE PROGRAM AS PERMITTED ABOVE, BE LIABLE TO YOU FOR DAMAGES, INCLUDING ANY GENERAL, SPECIAL, INCIDENTAL OR CONSEQUENTIAL DAMAGES ARISING OUT OF THE USE OR INABILITY TO USE THE PROGRAM (INCLUDING BUT NOT LIMITED TO LOSS OF DATA OR DATA BEING RENDERED INACCURATE OR LOSSES SUSTAINED BY YOU OR THIRD PARTIES OR A FAILURE OF THE PROGRAM TO OPERATE WITH ANY OTHER PROGRAMS), EVEN IF SUCH HOLDER OR OTHER PARTY HAS BEEN ADVISED OF THE POSSIBILITY OF SUCH DAMAGES.

Index