

task-guidelines-and-examples11.txt
LOGCLEF'2009: Log Analysis and Geographic Query Identification (LAGI)
Task Guidelines and Examples (version 2.1)

Alexander Yeh, Inderjeet Mani, Christine Doran

July 31, 2009

Contact email: asy at mitre.org (' at ' -> '@')

Copyright 2009 The MITRE Corporation. All rights reserved.
Approved for Public Release; Distribution Unlimited. Case# 09-3188

=====

The task is to identify geographic elements in search log queries.

This is being done for two sets of logs:

1. Tumba! - a Portuguese web search engine
2. The European Library (TEL) -
on line search for materials in various libraries in Europe.
We are looking at the subset in English (which are the majority of
the logs).

=====

2009 Schedule

Time zone for times: Eastern United States Daylight Savings time

with this version of the guidelines and examples,
small sample data files are available, both before and after annotation.
Unfortunately, there is no time for producing larger training data sets.

July 31: Test data will be available at
<http://app01.iw.uni-hildesheim.de/~clef/LogCLEF/>

August 7: The annotated test data file are due back to Alexander Yeh by midnight.
They can be sent by email to: asy at mitre.org (' at ' -> '@').
The test data will have an UTF8 character encoding.
To try to preserve this encoding and the line endings used in the data
files,
please send the files back as attachments
that have been compressed with gzip
(some compression schemes may change the line endings used in the text).
Please send the Tumba! and TEL data files back as separate files.

August 14: The results will be sent to the participants.
Recall, precision and balanced F scores will be calculated
for finding places in the data sets.

Note that for various reasons,
it is possible that we will not score some parts of the test data.
If we do this, we will let you know what parts were not scored
when we send out the results.

=====

Rules:

0. The TEL queries have a large overlap with the dataset in
Log Analysis for Digital Societies (LADS),
the other track in LogCLEF 2009.

You are NOT allowed to use the LADS dataset to help you with the LAGI task.

1. A query is a geographical query if and only if it is bounded geographically.
Thus, the queries "restaurant" or "restaurant food" are not considered geographical queries.
2. The task is to mark the query with zero or more non-overlapping place tags indicating geographical elements. Each place tag marks a substring (proper or not) of the query, which we will call a place term. Queries that are classified as not being a geographical query are not marked with any place tags.
3. A place term can be any country, a city or town, geographical feature, building, stadium, university, store, restaurant, statue, etc. described as a place in wikipedia (Portuguese wikipedia for Tumba! and English wikipedia for our English subset of TEL). Wikipedia was chosen because it is reasonably comprehensive and is readily available. Since wikipedia is constantly being updated, the evaluation will be using the particular versions that are mentioned later on. While those versions will be used to generate the official answers, participants are also free to use other wikipedia resources as well.
4. A candidate place term can map to more than one possible meaning in wikipedia.
If the look-up of a candidate place term returns an initial wikipedia disambiguation page linking directly to at least one disambiguated page describing a place, the candidate place term is tagged as a place.
If the initial page from wikipedia is not a disambiguation page, but contains a link to a disambiguation page, the candidate place term is marked as a place only if the initial page describes a place. The initial page is taken to represent the predominant sense of the candidate term.
To do a look-up, type the candidate place term into the 'search' (English) / 'busca' (Portuguese) area and then either press the 'Enter' key or click 'Go' / 'Ir' (do not click 'Search' / 'Pesquisa').

This method of disambiguation is used when a query has no indicated preference for which sense to use. But if the query indicates a preference for a sense, then that sense is what is used. An example is the query 'casanova commune'.
A search for 'casanova commune' in the English wikipedia does not return an article. Rather it returns a 'search' page (instead of being an article for some term, the page gives a ranked list of articles that contain parts of the candidate place term somewhere in the articles' text), which is ignored in this evaluation.
For 'casanova', the English wikipedia returns an article on a person named Casanova, so that is the default predominant sense. But that article has a link to a disambiguation page, and the disambiguation page has a link to the place 'Casanova, Haute-Corse', which is a commune.
This query indicates that this sense of 'casanova' is the preferred one for the query, so this overrides the default preferred sense based on the initial page returned by the wikipedia.
5. A place term can occur in a title (of a book, movie, team, etc.), but the title itself (if a different text span from the place) is not to be tagged.
If the wikipedia being used mentions that the phrase being examined

is the title (of a book, movie, etc.), this wikipedia entry is ignored for the purposes of this evaluation. For example, if TEL had the query 'paris match', a look-up in the English wikipedia will return an article about a weekly magazine. For this evaluation, ignore the fact that 'paris match' is a weekly magazine.

The reason for this rule is to make query processing more interesting for this evaluation given that the TEL queries are full of titles of books, etc.

6. Capitalization (upper and lower case) in the query is ignored, as it is used inconsistently in the queries. Thus queries are treated as folded to lowercase.
7. Acronyms are treated like other potential place terms. Example acronyms include 'USA' for 'United States of America' and 'FCUL' for 'Faculdade de Ciências da Universidade de Lisboa'. Acronyms are searched for in the wikipedia as is, and are NOT expanded first. So for 'FCUL', one searches for 'FCUL' and not 'Faculdade de Ciências da Universidade de Lisboa'.
8. There are no embedded place tags in this evaluation. Only the largest extent of a place term is marked. So with 'South America', mark 'South America' as a place, and not 'America'. A place term can include qualifications that provide additional context. Thus, in 'cavan county ireland', only 'cavan county ireland' is marked.
9. wildcards ('*') are ignored. So 'iceland*' is treated as if it were 'iceland', for which the English wikipedia returns an article on the country 'Iceland'. 'ice*' is treated as if it were 'ice' (even though '*' will match 'land' and so 'ice*' will match 'iceland'), for which the English wikipedia returns an article about frozen water.
10. If some words of a query can be interpreted as forming a phrase, this will be preferred over interpreting those words as isolated words put in the same query. For example, with the query 'burlington university', given that the English wikipedia does not have an article on a place with that name, the preference is interpret this as a 2 word phrase rather than just the isolated words 'university' and 'burlington'. But with 'university burlington', for which the English wikipedia also does not have an article on a place with that name, one cannot interpret 'university burlington' as a phrase, so just interpret 'university burlington' as 2 isolated words. Also see Rule 11 below.
11. When some part of a query forms a phrase, the words in the phrase that are not nouns may NOT have their regular 'non-name' meaning given in wikipedia, and the wikipedia may list some possible name meanings for those words. In this case, one should ignore the wikipedia and use the 'regular non-name' meanings for these words if it makes sense.

Some examples (more details given later on):

'TEL: restaurant near university' - interpret 'near' as meaning 'close to'.

This meaning is not in the English wikipedia article on 'near'.

That article does mention that 'near' may refer to the place 'near east'.

This wikipedia interpretation of 'near' will be ignored.

task-guidelines-and-examples11.txt

'TEL: strongholds in xv century' - interpret 'in' as a preposition.
This meaning is not in the English wikipedia article on 'in'
(Not in the version being used.

The current English wikipedia does mention this meaning).
That article does mention 'in' as possibly referring to the
following places:

- * India (country code),
 - * Indiana in the US (postal abbreviation),
 - * Ingolstadt in Germany (something to do with cars).
- These place interpretations will be ignored.

'Tumba!: jornais de leiria' [in English: 'periodicals of leiria']
- interpret 'de' as a preposition ('of').

This meaning is not in the Portuguese wikipedia article on 'de'.
That article does mention that 'de' may refer to the
following places:

- * Germany (Deutschland),
- * Delaware.

These place interpretations will be ignored.

12. Places that have never existed (imaginary places) are not marked.

An example of such a place: 'Wonderland' in the story
"Alice's Adventures in Wonderland".

13. Redirection: when there are many ways to refer to a topic, a wikipedia
will often put an article on the topic just under one way to refer to it.
The other ways will each just have a short 'article' that is a redirection
page: a page that just points to the one way of reference that has the
actual article. Usually, a wikipedia automatically follows the links in
such redirection pages and you do not have to do anything.

For example, in English, a look-up of 'sicilia'
(using the 2008 dumps mentioned below):

<http://app01.iw.uni-hildesheim.de/wikiclef/wiki-en/index.php5/Sicilia>
returns an article on 'Sicily' with the note '(Redirected from Sicilia)'.
Another example is in Portuguese, a look-up of 'marinha grande':

http://app01.iw.uni-hildesheim.de/wikiclef/wiki-pt/index.php5/Marinha_grande
returns an article on 'Marinha Grande' with the note
'(Redirecionado de Marinha grande)'.
If the wikipedia does not automatically follow-up such links,

then you should follow these links yourself and act as if the wikipedia
did automatically follow these links.

An example is in the disambiguation article on 'parisian':

[http://app01.iw.uni-hildesheim.de/wikiclef/wiki-en/index.php5/Parisian_\(disambiguati](http://app01.iw.uni-hildesheim.de/wikiclef/wiki-en/index.php5/Parisian_(disambiguati%20on))
on)

the link to 'Parisian (person)' is the page:

[http://app01.iw.uni-hildesheim.de/wikiclef/wiki-en/index.php5/Parisian_\(person\)](http://app01.iw.uni-hildesheim.de/wikiclef/wiki-en/index.php5/Parisian_(person))
This page is a 'Redirect page' to 'Paris' (following the link leads to an
article on 'Paris').

In this case, act as if the page on 'Parisian_(person)' is the article
on 'Paris'.

=====

wikipedia

As mentioned in the rules above, this task makes use of the wikipedia.
The task will make use of read-only restored versions of the
wikipedia from 2008 dumps,

25th of June for Portuguese:

http://app01.iw.uni-hildesheim.de/wikiclef/wiki-pt/index.php5/P%C3%A1gina_principal

task-guidelines-and-examples11.txt

and 24th of May for English:

http://app01.iw.uni-hildesheim.de/wikiclef/wiki-en/index.php5/Main_Page

The Portuguese wikipedia is quite large and the English wikipedia is even larger. It is possible that it will be somewhat slow when multiple groups request look-ups at roughly the same time.

Under heavy load conditions, we have sometimes observed some wikipedia servers either return an empty web-page or time-out trying to respond.

If you will look-up the same term multiple times, it may help you to store the results of your first look-up of a term so that you can refer to those results locally when you want to look-up the same term again later on.

Also, for faster responses, you may wish to use another version of the wikipedias on other web-sites (including the live wikipedias en.wikipedia.org and pt.wikipedia.org).

Be advised that these other versions will give responses that may be different from the 2008 versions mentioned above,

and we used the above 2008 versions to generate the answer key.

In addition, due to differing amounts of network congestion and other factors,

* The response time may vary drastically at different times.

For example, using the same set of look-ups on en.wikipedia.org multiple times produced response times ranging from 19 seconds to 109 seconds.

* When comparing different sites (for example, en.wikipedia.org and the restored 2008 English dump mentioned above), which site is faster can depend on both the phrase being looked-up and when the look-up is attempted.

Those pages for the 2008 dumps also give links to the dumps used in the restoration for those interested

(the dump for English comes in 2 parts, which need to be concatenated together using something like 'cat' in Unix).

As a warning, re-populating a wikipedia with these dumps is hard.

Because of the difficulty in restoring the wikipedias, the version of the English wikipedia being used in this evaluation

has two restrictions on it:

A. It has not been indexed for searching in the text of the articles (matching a term to an article title still occurs).

This will not matter for the evaluation because as mentioned in rule 4, the results found in 'search' page are ignored in this evaluation.

B. It does not contain articles with titles that have 'special' characters. For example, the articles on "André Félibien", "Ångström" and "Crécy-en-Ponthieu" are missing.

Because of this, when dealing with English, the evaluation will not look at data containing special characters.

Note that these two restrictions are for the version of the English wikipedia being used.

The version of the Portuguese wikipedia being used does not have these restrictions (although with rule 4, the search results returned by the Portuguese wikipedia are ignored).

An alternative url to reach both wikipedias: <http://www.uni-hildesheim.de/logclef/>

The following may be helpful.

The files [ptwiki-20080625pages-articles-titles.txt.gz](#)

and [enwiki-20080524pages-articles-titles.txt.gz](#)

are available at the web-site <http://app01.iw.uni-hildesheim.de/~clef/LogCLEF/>

These files contain the lines with the "<title>...</title>" fields

from the 2008 dumps.

These fields are the titles of the articles, so these files contain the

titles of all the articles/pages in the dumps.

This may be useful to quickly determine whether there is an article/page with a particular name.

task-guidelines-and-examples11.txt

=====

Examples:

Note: the format shown in the examples below is a preliminary format.
The exact format for the results and scoring method will be posted later.

Tumba!: jornais de leiria [in English: 'periodicals of leiria' (In Portugal, there is among other places, a municipality, district and an urban community with the name Leiria)]

Tagging:

jornais de <place>leiria</place>

#'Leiria' can refer to more than one possible place (possibilities
#listed in the Leiria disambiguation page in the Portuguese wikipedia).
#But because 'Leiria' is a place name, it still gets marked as a place
#(unlike the 'restaurant' example above).

Tumba!: chaves

Tagging:

<place>chaves</place>

#An ambiguous case. A Portuguese wikipedia look-up of 'chaves'
#returns a disambiguation page,
#which lists as possible meanings the name of a number of different
#places, as well as other meanings
#(including a television show and a character in that show).
#For this evaluation, the emphasis is on possible places, especially
#names of places,
#so when such an ambiguity is found, mark it as a place.

TEL: central europe

Tagging:

<place>central europe</place>

#A look-up in the English wikipedia of 'central europe' returns an
#article describing a place.
#So in this evaluation, 'central europe' counts as a place, even if
#its true status as a place is fuzzy.
#And because 'central europe' gets marked as a place, 'europe' does
#not (no embedded markup) nor does 'central' (looking-up 'central'
#returns a page with many possible meanings, including several named
#places, but again, no embedded markup).

TEL: sicilia

Tagging:

<place>sicilia</place>

#'sicilia' is Sicilian for the place called 'Sicily' in English.
#But the English wikipedia returns an article describing a place when
#given 'sicilia', so this will count as a place.
#On a related note, a look-up of 'sicilia' in wikipedia returns a
#redirection page which points to the article on Sicily, but this will
#still count as a place.

Tumba!: origem do vidro na marinha grande [in English, possibilities
include something like: 'origin of glass in large navy' or 'origin of
glass in Marinha Grande' ('Marinha Grande' is a city in Portugal)]

Tagging:

origem do vidro na <place>marinha grande</place>

#In the Portuguese wikipedia, looking up 'marinha' returns an article
#on 'navy' (and related topics like fishing fleets and merchant ships),

task-guidelines-and-examples11.txt

#but looking up 'marinha grande' returns an article on a city.
#The larger construct refers to a city, so that interpretation is used.

TEL: casanova

No tags.

#The wikipedia often returns possible places for words that
#predominantly have a non-place meaning.

#This is one of those cases.

#For 'casanova', the English wikipedia returns an article (in this
#case via redirection) on a person 'Giacomo Casanova'.

#That article has a link to a disambiguation page which lists a number
#of possible meanings for 'casanova', including the place

##'Casanova, Haute-Corse'.

#To handle such situations, we will use the following rule:

If there is a disambiguation page that gives both possible place
and non-place meanings for a look-up,

but the initial article returned for the term gives a non-place
meaning and it is via this article that one gets a link to the

disambiguation page,

then the term will be deemed to predominantly have a non-place meaning

and be deemed not a place.

#Side note: there is a

##'http://app01.iw.uni-hildesheim.de/wikiclef/wiki-en/index.php5/Casanova,_Virginia'

#article in the wikipedia, but it is not connected to

#the 'casanova' disambiguation page.

TEL: casanova commune

Tagging:

<place>casanova commune</place>

#So far, the we have mentioned the use of the wikipedia's ordering of
#preferences to disambiguate places from non-places.

#When a query itself gives enough context to indicate a desired sense,
#then this sense overrides the wikipedia's ordering of preferences.

#In this example, the query indicated that the 'Haute-Corse commune'

#sense of 'casanova' in the wikipedia is desired, so this overrides

#wikipedia's own ordering (via initial page returned).

Tumba!: escolas de marinho [in English, possibilities include:
'schools of sailing' ('sailing schools') or 'schools of Marinhoiro'
(Marinhoiro is a place in Brazil)]

No tags.

##'marinhoiro' means sailor in Portuguese.

#It is also the name of a place in Brazil.

#In the Portuguese wikipedia, a look-up of 'marinhoiro' returns an

#article in sailors/sailing,

#not a place in Brazil.

#That article links to a disambiguation page that mentions some other

#non-place meanings, but these are ignored and even if not ignored,

#would not change the results.

TEL: strongholds in xv century

No tags.

##'strongholds' is not bounded geographically and not the name of a
#place.

##'strongholds' is has a common (non-name) meaning and is not in the
#English wikipedia (no article exists).

##'stronghold' is in the English wikipedia and some of the possible
#meanings are names of places.

#For 2 reasons, the place name interpretation is being ignored (either

#reason is sufficient):

1. Looking for 'stronghold' in the English wikipedia returns an
article about fortified places in general, and not a particular
bounded place, nor places with a particular name.
This article has a link to a disambiguation page which lists a
number of possible meanings for 'stronghold', including being the
name of some places.
So like the 'casanova' example above, the predominant
interpretation of 'stronghold' will be that of the initially
found article, which in this case is not a specific bounded
place, nor places with a specific name.
2. 'strongholds' is the plural of 'stronghold'.
while 'strongholds' could refer to a group of places, all with
the name of 'stronghold', this possibility is being ignored in
this evaluation unless the plural form itself is considered a
name or the appropriate wikipedia has an article about this
possibility.

TEL: iceland*

Tagging:

<place>iceland</place>*

#The English wikipedia returns an article on the country of Iceland
#when given this term.
#This article in turn links to a disambiguation article for 'iceland',
#which has various place and non-place interpretations.
#So the 'country' interpretation is the predominant interpretation.

TEL: ice*

No tags.

#The '*' operator can match anything (so '*' can match 'land' and 'ice*'
#can match 'iceland'), but as it written:
#For 'ice', the English wikipedia returns an article about frozen water.
#This article points to a disambiguation article, but like the
'casanova' and 'stronghold' examples above, the predominant sense is
#considered to be that of the article initially found.

TEL: cavan county ireland 1870

Tagging:

<place>cavan county ireland</place> 1870

#Note that the article name does not have to exactly match the
'word/phrase' to count as a match (the article on 'County Cavan' is
#deemed to match the phrase 'cavan county ireland', with the
#appropriate page found via a disambiguation page on 'cavan').

#Looking-up 'ireland' in the English wikipedia returns an article on
#the island 'ireland', a place.
#This article then has pointers to other definitions of 'ireland',
#both places and non-places.
#Looking-up 'cavan county' in that wikipedia returns a search page
#which finds articles with mentions of 'cavan county'.
#I ignored these (see the following example for an more details on why
#they are being ignored).
#Looking-up 'cavan' in the wikipedia, it returns an article on a town
#named 'Cavan' in Ireland.
#This article points to a disambiguation page.
#Since the query is looking for a county, and not a town, I look in
#the disambiguation page.
#This page points to an article on 'County Cavan'.
#The construction of the query suggests this interpretation is what is
#meant. So even though the default predominant interpretation of

task-guidelines-and-examples11.txt

#'cavan' is a town, according to the wikipedia, the query's
#construction indicates that a county is called for.

#Also, looking for 'county' in the wikipedia returns its 'usual'
#meaning plus a few variants of that meaning.

Tumba!: saksa

No tags.

#A Portuguese wikipedia look-up of 'saksa' finds no article with that
#name and returns a search page.

#'saksa' is Germany in Finnish and some of the results of that
#search page indicate this.

#But for this evaluation, this will not count, as the initial page is
#neither a page describing a place nor a disambiguation page linked to
#at least one disambiguated page describing a place.

#In any case, the restore of the English wikipedia dump used

#in this evaluation has not been indexed for search and the search pages
#from using this dump will not have any results.

TEL: university

No tags.

#An English wikipedia look-up returns an article on 'university' as a
#place for higher education (this article has a link to a
#disambiguation page with other possible meanings for 'university',
#but these other meanings are being ignored as the non-predominant
#meanings).

#This place is not bounded geographically, so no tags.

TEL: universities

No tags.

#An English wikipedia look-up returns a redirection to an article on
'university' as a place for higher education.

#This place is not bounded geographically, so no tags.

TEL: restaurant near university

No tags.

#An English wikipedia look-up of 'university' is described above.

#A similar look-up of 'restaurant' returns an article on 'restaurant'
#as a place for people to be served food and drink (this article has a
#link to a disambiguation page with other possible meanings for
'restaurant', but these other meanings are being ignored as the
#non-predominant meanings).

#This place is also not bounded geographically.

#So this query is not bounded geographically.

#It could involve all universities on earth and all restaurants near
#them. So no tags.

TEL: universities burlington

Tagging:

<place>universities</place> <place>burlington</place>

#A look-up of 'burlington' in the English wikipedia returns a
#disambiguation page listing numerous interpretations, including the
#names of many places. So 'burlington' will be annotated as a place.
#'universities' is now restricted geographically (within some place
#named 'burlington'), so it is also annotated as a place.

#Note that this mark-up is occurring even though the phrase is
#describing multiple universities in multiple places named 'burlington'.

task-guidelines-and-examples11.txt

TEL: university burlington

Tagging:

<place>university</place> <place>burlington</place>
#Similar to the 'universities burlington' example above

TEL: burlington's universities

Tagging:

<place>burlington</place>'s <place>universities</place>
#Similar to the 'universities burlington' example above

TEL: burlington university

Tagging:

burlington <place>university</place>
#The difference between this and 'university burlington' is that this
#is being treated as a phrase, where 'burlington' is being used in a
#predicative sense.
#This is more obvious in the phrase 'brazilian university' (which
#would get annotated as 'brazilian <place>university</place>'), where
#the word to the left of 'university' is obviously an adjective and
#not a noun.
#Also, 'burlington university' and 'brazilian university' are not
#actual names of places according to the wikipedia.
#Looking them up in the wikipedia just returns a search page, not an
#article describing them.
#So even though 'burlington university' and 'brazilian university' in
#some sense seem to actually exist
#(http://realtlinkdev.com/properties/view/Burlington_University/50
#and http://www.wipo.int/sme/en/best_practices/unicamp.htm ,
#respectively),
#they are being ignored because the wikipedia has no article on them.
'burlington university' and 'brazilian university' have some
#place annotations because they are about places.
#Something like 'burlington book' or 'brazilian book' will have no
#place annotations at all because a 'book' is not a place of any sort.

TEL: burlington universities

Tagging:

burlington <place>universities</place>
#Similar to 'burlington university'

TEL: suffolk university

Tagging:

<place>suffolk university</place>
#Looking up 'suffolk' in the English wikipedia returns an article on a
#place (a county in England). This article links to a disambiguation
#article that links to a number of alternative meanings, including the
#names of other places.
#The difference between 'suffolk university' and 'burlington university'
#(or 'brazilian university') is that 'suffolk university' is the proper
#name of some university (as opposed to 'burlington university',
#which refers to any university associated with burlington).
#A look-up of 'suffolk university' in the wikipedia returns an article
#on a university with that name.
#A similar look-up of 'burlington university' or
#'brazilian university' returns no article, just a search page.
#So the entire phrase is marked as one name.
#There are no embedded tags in this evaluation, so 'suffolk' and
#'university' are NOT annotated as places by themselves in the example.

task-guidelines-and-examples11.txt

TEL: university of burlington

Tagging:

<place>university</place> of <place>burlington</place>

#Similar to the 'university burlington' example.

#'university of burlington' is not the actual name of a university

#according to the English wikipedia:

#A look-up of this phrase returns just a search page, not a article of

#a university with this name.

TEL: universities of burlington

Tagging:

<place>universities</place> of <place>burlington</place>

#Similar to the 'university of burlington' example.

TEL: university of lisbon

Tagging:

<place>university of lisbon</place>

#Similar to the 'suffolk university' example.

#The example here is the proper name of an actual university with an

#article in the English wikipedia.

#Like with the 'suffolk university' example, even though 'lisbon' by

#itself would be annotated (and hence also 'university'), there are no

#embedded tags, so just the outermost extent (with the entire name) is

#annotated.

Tumba! Query Format

Each line in a data file are the words for one query
(in the same word order as in the query).

There may be some '+'s and/or '"' (double quote marks).

Treat a '+' like a space.

Ignore a '"'.

In front of each query are two numbers and two '@' characters in the
sequence '[number] @ [number] @'

Just ignore (neither alter nor annotate) this preliminary character sequence.

Examples include:

4333825 @ 4777 @ "administração escolar"

4933229 @ 7888 @ "escola+hip+hop"

6971716 @ 106342 @ jornais de leiria

which give, respectively, the queries:

"administração escolar"

"escola+hip+hop"

jornais de leiria

European Library Query Formats

A. A query string may appear enclosed inside a set of "'s.
An example, "university of burlington"

B. In that string, a space may be replaced by a '+'.
An example: "university+of+burlington"

- C. A string may be placed inside a set of parentheses.
One example: ("university of burlington")
Another example: ("university+of+burlington")
- D. Inside the parentheses, just before the string, there may be one or two words to indicate the field in the catalog and the type of search to do with the string inside the parentheses.
Two examples: (title all "university of burlington")
(title all "university+of+burlington")
which indicates to look for a title containing the words in the string.
The first word before the string can be one of the following:
title
creator
subject
type
language
isbn
issn
publisher
The second word is 'all' (use all the words in the search) or 'exact' (only match the exact phrase in the search).
Even with 'all', for this evaluation, rule 10 still applies: interpret the words in the string as a phrase when possible, as opposed to the words in the string being isolated words.

Please ignore the text for specifying the 'language' field.
For example,
in '(language all "eng")',
ignore "eng", which stands for English
(the English wikipedia mentions 'English', 'England' and some other possible meanings for "eng").
- E. These sets of parentheses may be combined with the boolean operators 'and', 'or' and 'not'.
In this evaluation, the only boolean operator that will be present is 'and',
which will indicate that all the groups of words in the strings being present in a search item is the most desirable situation.
Rule 10 does not apply to words from different strings being combined together with 'and'. So for example, in the query

("burlington") and ("university")

treat "burlington" and "university" as isolated words and not the phrase 'burlington university',
even though rule 10 would treat the string "burlington university" as a 2 word phrase.
- F. A variation of format A: no enclosing "'s.
An example, university of burlington
- G. A variation of format D: no enclosing (or).
An example: title all "university of burlington"

Note that the actual TEL queries may include some that are combinations of the above formats.

For example: burlington and ("university")
These will not be included in this evaluation.

Each query is on one line.
Similar to Tumba! (but with a '&' instead of a '@'),

task-guidelines-and-examples11.txt

in front of each query are two numbers
and two '&' characters in the sequence '[number] & [number] &'
Just ignore (neither alter nor annotate) this preliminary character sequence.

Examples include:

902980 & 482 & (creator all "casanova")

906474 & 15432 & casanova

712725 & 5409 & ("cavan county ireland 1870")

which give, respectively, the queries:

(creator all "casanova")

casanova

("cavan county ireland 1870")

=====