

Searching through the materia medica

Ing. Peter Bezemek

Affiliations of the author: AEON GROUP Ltd. co-owner and director
Correspondence and reprints: AEON GROUP Ltd., Tr. SNP 39, 974 01 Banska Bystrica,
Slovakia;
email < support@mercurius.sk >; web: < <http://www.mercurius.sk> >

BANSKA BYSTRICA, SLOVAKIA, July 2010

Abstract

The document describes the advanced searching techniques for the materia medica texts, as implemented in Mercurius homeopathic software (Materia Medica Explorer).

Introduction

Considering the fact that all current and old repertories contain many rubrics which are incomplete (i.e. they do not contain all remedies they should), effective search of materia medica (which is the primary source upon which to base the homeopathic prescription) would allow a homeopath to utilize this information. But even with perfect repertory rubrics, materia medica provides the necessary context and specifics not present in the generalized (and therefore imprecise, to a degree) rubric of the repertory.

However, when it comes to looking up specific symptoms, materia medica (as opposed to the repertory) present many challenges which render the task quite difficult. Various authors use different words and word sequences to describe essentially the same phenomenon which often makes locating the matching parts very troublesome.

The document describes the effective searching techniques as implemented in Mercurius homeopathic software (Materia Medica Explorer).

Consideration of the specifics of materia medica search compared to repertory search

As mentioned in the introduction, materia medica search provides certain challenges when it comes performing it efficiently and with the aim of obtaining high-quality results.

While the search processes as described in the previous related document (Searching through the text-based homeopathic repertory) have proved effective and sufficient for the repertory use, the application of similar algorithms to materia medica search has been less effective, mainly due to lack of structuring that is inherent to repertory.

Three most frequent problems connected with materia medica search (as we perceive them):

1. When searching with basic keywords method, there are sometimes erroneous results, especially in materia medicae that tend to use long sentences. Let's consider an example - we wish to find remedies that have symptom of "headache ameliorated by eating". Searching for "Headache Eating Amel" returns many useful results, but also results which are clearly not useful, such as this entry from Hering's Lachesis:

*"Toothache: from **eating**; biting; something warm; drinking warm things or cold; during menses; when awaking; from getting wet; in the Spring; in Summer; **amel.** from external warmth; from cleaning teeth; with chilliness, heat, thirst; in persons who have taken too much mercury; with **headache**; with swelling of cheek; teeth too long, blunt."*

This is a case of a very long sentence which contains the necessary keywords, but they are quite far from each other which significantly reduces a probability of useful result. This problem could be remedied if we could specify a maximum distance between the words "eating" and "amelioration" which are very likely to be quite close to one another in a useful match.

Examples of useful matches which demonstrate the concept of "useful nearness":

"*headache amel.* while *eating*"

"*Headache* in *sinciput* after being kept waiting for dinner; **amel. after *eating*"**

*"severe **headaches**, with sensation of weight at back of eyeballs, and a feeling as though eyes would be drawn back into head; stiffness of nape of neck; on rising from seat great difficulty from a feeling of stiffness of limbs; pain in bottom of back, around waist, also around shoulders and nape of neck; numbness of right arm and hand; burning sensation from pit of stomach to throat and ears, **amel.** by *eating*;"*

"*Headache*, in morning, on rising, **amel. *eating*;"**

"*Headache* before and during menses; right eye to occiput, **amel. heat, pressure, lying down, *eating*, gentle motion;"**

"Eating amel. headache; pain in eyes; nausea; cough."

To prevent this situation, we can define a NEAR operator which sets a maximum allowed distance between the words for sentence/paragraph to count as a match.

2. When using synonyms during the search (which is almost always necessary when searching through MM), the format of synonyms sometimes caused an overabundance of results due to very broad matches on certain terms. For example, searching for „Alcohol“ would also produce matches containing the word „drunk“, which could be related, but also may be overly broad match for the momentary purpose. This problem can be resolved by better user control of how synonyms are utilized during a particular search i.e. various levels of synonyms are necessary.
3. Especially (but not exclusively) in cases when a search query includes certain anatomical parts (for example, we are looking for “headache in morning”), there are many instances of valid matches that cannot be found using the standard search algorithm. For example “stitching in the temples in morning” would not be found, although it is clearly a valid match for our query. This can only be resolved by a radically different and more advanced implementation of synonyms. Please note that this problem is not present in repertories which would include the symptom in a rubric such as “Head, pain, temples, stitching, morning”.

Our new searching engine effectively addresses all these situations.

Basic search

Since we use text-based source, we need to search for text.

Each sentence/paragraph of the materia medica can be parsed into words which can be searched for.

In Mercurius, all you need to do to find a sentence/paragraph match is typing the **words or parts of the words** you think should be present in the sentence/paragraph, in any order.

Full Partial matching

To search for a pain in the temples that moves from left to the right temple, we could type the search query

*Query 1: **temple left right pain***

The query executes as searching for rubrics containing the characters “*temple*” AND “*left*” AND “*right*” AND “*pain*” (according to star convention¹)

¹ A mask for representing the names of multiple entries with asterisks. Asterisk (*) means any characters i.e. “*head*” matches the words “fore**head**” or “**head**ache”.

Some of the results triggered by the query (the search results are from repertory search but it works in the same way in MM):

Head, phenomena, **pain**, drawing, forehead, sides, **right**, **left** and to **left temple**, to
Head, phenomena, **pain**, **temples**, **left**, **right**, to
Head, phenomena, **pain**, bursting, **temples**, **left**, **right**, then
Head, phenomena, **pain**, neuralgic, **temples**, **left**, **right**, to

As it is clear from the results, it is important to be able to search not only for whole words, but also parts of the words, otherwise the keyword "temple" would not trigger the results that contain the word "temples".

This is a very important feature, useful for finding the words about which we are unsure in what form they are used in the rubric/symptom. It is especially useful in some non-English languages such as German or Slavonic languages which use a lot prepositions and different word forms, making it very difficult to find anything using plain whole-word matching.

Full partial matching avoids this problem and allows the user to find the matching rubrics without the need to think about all the possible word forms.

One-sided partial matching or whole-word matching

Sometimes, full partial matching may trigger undesired results, such as when searching for "head" finds rubrics with "forehead".

This can be avoided by using one-sided partial matching or whole-word matching. Such behaviour is set by using the capital letters at the beginning or the end of the keyword.

Rules:

1. Left-sided partial matching - > If the keyword starts with a capital letter, the matching word starts with the keyword . ("**Head**" triggers "**head**" or "**head**ache", but not "fore**head**")
2. Right-sided partial matching - > If the keyword ends with a capital letter, the matching word ends with the keyword.
("**head**" triggers "**head**", "fore**head**", but not "**head**ache")
3. Whole-word matching - > If the keyword starts and ends with a capital letter, the matching word must match the keyword exactly. ("**Head**" or "**HeAD**" or "**HEAD**", all trigger "**head**" and nothing else)

For English language, the most suitable search seems to be left-sided partial matching.

Negative matching

To specify a negative match, we can use the minus "-" sign to filter out certain keywords we do not wish to be included in the results.

To look for rubrics related to hay fever occurring in spring or summer, but not in autumn or winter, we can specify the following query:

*Query 2: "**hay fever -autumn -winter**"*

Some of the results triggered by the query:

*Nose, phenomena, coryza, general, annual, **hay fever***
*Nose, phenomena, coryza, general, annual, **hay fever**, spring*
*Nose, phenomena, coryza, general, annual, **hay fever**, summer*
*Nose, phenomena, coryza, general, annual, **hay fever**, summer, diarrhea, in*
*Nose, phenomena, coryza, general, annual, **hay fever**, august, in*

OR matching

It is sometimes useful to be able to use OR function to specify.

In Mercurius, OR matching is specified by plus "+" sign.

To look for the previous example using the OR matching, we could write.

*Query 3: "**hay fever spring + hay fever summer**"*

Some of the results triggered by the query:

*Nose, phenomena, coryza, general, annual, **hay fever, spring***
*Nose, phenomena, coryza, general, annual, **hay fever, summer***
*Nose, phenomena, coryza, general, annual, **hay fever, summer**, diarrhea, in*

NEAR matching

NEAR operator (character "/") is used to define a maximum number of words that may occur between two search terms for sentence/paragraph to count as a match.

Syntax: <word1> /x1 <word2> /x2 <word3>...

where x1, x2 are allowed distances. If no number is specified, default NEAR distance is used.

Example: "*Eating /1 Amel*" search for all occurrences of words starting with "eating" and "amel" (such as "ameliorates") which contain a maximum of one word between them.

Examples of matches:

*"The headaches are digging, throbbing in character, **amel.** by **eating**, but agg. by any attempt to use the mind or by overwork."*

*(Throat) "Dryness in forenoon, **amel. eating.**"*

*"Many complaints are **ameliorated** by **eating.**"*

One-sided NEAR matching

Same as NEAR matching, but also define the directionality (the order of the words).

Syntax for left-to-right matching: <word1> >/x1 <word2>

Syntax for right-to-left matching: <word1> </x1 <word2>

Example: "*Eating >/1 Amel*" search for all occurrences of words starting with "eating" and "amel" (such as "ameliorates") which contain a maximum of one word between them AND where "eating" precedes "amel". Note that none of previous example results matches the criteria because the order of the search words is reversed. If we used "*Eating </1 Amel*" search query instead, they would all match.

Negative NEAR matching

Negative NEAR matching is also allowed and is useful for instances when you want certain words NOT to be near one another.

Syntax: <word1> /x1 -<word2>...

Example: "*Eating /3 -Amel Amel*" or "*Amel Eating /3 -Amel*" search for all occurrences of words starting with "eating" and "amel" (such as "ameliorates") which are more than 3 words apart (i.e. they are not within 3 words from one another). Note that without the NEAR operator, the search query would produce no results because it contains both a match and a negative match for the same word.

Synonymic matching

The support of synonyms is vital for achieving the best results. English expressions have lots of synonyms and many of them are used in materia medica.

By using synonymic matching the number of results can rise even 20 times!

Mercurius implements the best-in-the-industry support of synonyms.

Compared to previous versions, new Materia Medica Explorer features extremely advanced search engine developed specially for working with materia medica texts, which supports six levels of synonyms, which are completely user-defined and editable via one simple text file (mmsynonyms.cfg).

Moreover, the synonyms are now linked together (in a similar way as hyperlinks are used on the Web) so entire structures can be covered easily. This feature, in connection with various levels of synonyms can be extremely useful when searching for anatomical structures (such as "head") while only a subpart is present in the text (such as "temples").

Six levels of synonyms are defined like this:

Level 1 - "true" synonyms, also involving different forms of the same word, also singular/plural form (example: WINDPIPE triggers TRACHEA; AURICLE triggers AURICULA, PINNA, CONCHA, OUTER EAR, EXTERNAL EAR)

Level 2 - part of/kind of top-down relationship (example: EAR triggers OUTER EAR and MIDDLE EAR and INNER EAR, which in turn trigger MEATUS, LOBULE, AURICLE, COCHLEA, EUSTACHIAN etc. which then trigger other parts and subparts of ear)

Level 3 - pathological states (these are reserved for pathological states related to the trigger keyword, for example EAR triggers things like OTITIS, OTALGIA, TINNITUS, TYMPANITIS etc.)

Level 4 - broader synonym matches (example: EYE triggers EYEBROW or EYELASH; SINUS triggers UPPER JAW or CHEEK; NOSE ITCHING triggers HAY FEVER or RHINITIS; ALCOHOL triggers DRUNK)

Level 5 - even broader synonym matches (example: EYE triggers VISION; BREATHING triggers RALES; ASHAMED triggers DISCOURAGED or DISHEARTENED)

Level 6 - bottom-top follow link - it will only be followed ONE level up, not more (example: CAT triggers ANIMAL; OROPHARYNX triggers THROAT)

Multiple-words synonymic matching

Mercurius is the only known homeopathic software in the world that implements multiple words synonymic matching for materia medica.

This feature is extremely useful for triggering synonyms based on a combination of keywords.

To look for a headache with the perspiration of hands you would type:

Query 4: "**head pain perspir hand**"

or

Query 4: "**headache perspir hand**"

With synonymic matching, this would reveal the symptom:

"Sweaty hands and feet, during migraine."

As it is clear from the result, "head pain" combination propagated to "migraine" and discovered a symptom you would normally hardly find.

When the query is executed, the algorithm performs the search as if defined as:

"head pain perspir hand + headache perspir hand + migrai perspir hand"

The algorithm is able to map the one word or a multiple combination of words to one word or a multiple combination of words.

Thus "headache" triggers also "head pain".

By using these advantages, the following feats are also possible:

To look for a headache that gets worse outside, you just type:

Query 5: "**headache outside worse**"

which triggers the results, such as:

"Scalp sensitive to touch, roots of hair **hurt** when hair is moved; **agg.** when walking in **open air** and from a draught of air; amel. by a strong external **pressure** and scratching **head.**"

Please note that the synonym for search term "outside" is defined as "AIR /3 OPEN" so only occurrences of AIR and OPEN within 3 words apart would constitute a match.

Internally, the algorithm executes a complex query which covers the combination of all the defined synonyms.

Format of mmsynonyms.cfg

Synonymic engine loads the expressions from the file mmsynonyms.cfg, which can be edited by the user.

The synonymic combinations are listed line by line and are divided by semicolon. The first expression (called the trigger expression) is a word or a combination of words that triggers the rest of the expressions (called bullet expressions).

AURICLE;AURICULA;PINNA;CONCHA;OUTER /0 EAR;EXTERNAL >/0 EAR

In the example above "AURICLE" is the trigger expression, while the rest are bullet expressions. Please note that the trigger expression is also used as a bullet expression.

Both trigger and bullet expressions use the same syntax as is used during the search. In addition to this, there are several elements added which are not present in normal MM search.

The bullet expressions behave in accordance with the standard search used in Mercurius. Note that the trigger expression triggers the bullet expressions, but **not the other way round** (unless otherwise is indicated by use of | trigger mark).

For "respiration accelerated" to trigger "hyperventilation", it must be specified, for example:

Respir / Acceler;Hyperventila

This behaviour is very suitable for situations when you would like "eruption" to trigger all sorts of eruptions, but "herpes" to trigger only herpes.

Eruption;BOIL;BOILS;Carbuncle;Furuncle;Vesicle;Impetigo;Herpes;Blister;Pimple;Pustul;Papul;Nodul;Comedone;Acne;Condylom;SCALES;SCALY;Exanthem;Crust;Tubercle;Dan druf;Rash;Nettlerash;Urticar;Eczema;Blotch;Molluscum;Miliar;Wart;Scurf;Morphea;Enan them;Dermatitis;Caruncle

Beside these, the format uses several modifying elements:

1. trigger mark
2. trigger and bullet levels
3. follow mark
4. follow level constraints

Trigger mark

Unless trigger mark is used, only the first term (called "true trigger") in the line is considered a trigger expression. When trigger mark | is used before the search term, it also serves as a trigger. If trigger mark is used before the first term (true trigger), all bullet expression also serve as trigger expression.

So in

AURICLE;|AURICULA;|PINNA;|CONCHA;|OUTER /0 EAR;|EXTERNAL >/0 EAR

which is an equivalent of

|AURICLE;AURICULA;PINNA;CONCHA;OUTER /0 EAR;EXTERNAL >/0 EAR

all terms can serve either as triggers or bullets. Please note that NEAR operator is only used when the search term is used as a bullet; it does not have to be present in the search query for trigger to fire (so searching for OUTER EAR would trigger the response despite the fact the search query does not contain NEAR operator, but the synonym definition does).

Trigger and bullet levels

Materia Medica Explorer currently uses 6 levels of synonyms, as defined earlier. The level at which trigger or bullet expression is used is indicated by level or levels.

Syntax: [<level number><level number>...]

[2] means the definition is only valid if level 2 is switched on.

[12] means the definition is only valid if levels 1 and/or 2 are switched on.

[124] means the definition is only valid if levels 1 and/or 2 and/or 4 are switched on.

No level definition means any level will satisfy the level condition.

Using the following definition

AURICLE;|AURICULA;|PINNA;|CONCHA;|OUTER /0 EAR;[2]MEATUS;[2]LOBULE;
[2]AUDITORY /0 CANAL[1];[23]TEST_EXAMPLE

the word MEATUS would be used for trigger PINNA only if level 2 is switched on. The word TEST_EXAMPLE would be used for trigger AURICULA only if level 2 or 3 is switched on.

Follow mark

~ is a follow mark that indicates the term should be followed and used as a trigger in any matching line. This enables the creation of word nets which is necessary for covering anatomical relationships and other connections.

This example covers various anatomical parts of the ear.

EAR;|EARS;~[2]MIDDLE /0 EAR;~[2]OUTER /0 EAR;~[2]INNER /0 EAR;[5]~HEARING;
[4]AURAL;[5]MASTOID
INNER /0 EAR;[4]Vestibula;[2]COCHLEA;[2]MALLEUS;[2]Labyrinth
MIDDLE /0 EAR;[2]~EARDRUM;[2]Eustachian;[2]OSSICLE;[2]EAR[1]~ MEMBRANE
AURICLE;|AURICULA;|PINNA;|~CONCHA;|OUTER /0 EAR;|EXTERNAL[1]~ >/0
EARS[1]~;|~EARLOBE;~[2]MEATUS;~[2]LOBULE;[2]AUDITORY /0 CANAL[1]~;~[6]EAR
|CONCHA;CONCHAE;~[6]EAR

Note that part-of/kind of relationships are defined for level 2. When it is utilized, the search for a larger structure will also perform a search for all its subparts.

When the ~ mark is located at the start of the search term, the whole term is followed,

even if it consists of several words. If the follow mark is located at the end of multi-word search term, this word is followed separately.

This allows for trigger terms such as HEADACHE to propagate via HEAD~ / PAIN~ to include all parts of HEAD (forehead, temples, vertex, occiput etc.) and all kinds of PAIN (stitching, burning, aching, stinging, pressing, tearing etc.) towards many useful results.

Follow level constraints

These are the level constraints which define level to be used when following the bullet term.

For example

Headach;[2]Migrain;|HEAD[12]~ / PAIN[12]~

will follow HEAD and PAIN terms only with level 1 or 2, even if level 3 or higher is switched on by the user.