Information Engineering Group

MHB-F Specification

Document Revision 4

Andreas Seyfang
Katharina Kaiser

Vienna University of Technology
Institute of Software Technology & Interactive Systems
Information Engineering Group

Authors: **Andreas Seyfang**
             Katharina Kaiser

             seyfang@ifs.tuwien.ac.at
             http://ieg.tuwien.ac.at

Contact: **Vienna University of Technology**
             Institute of Software Technology and Interactive Systems

             Favoritenstraße 9-11/188
             A-1040 Vienna
             Austria, Europe

             Telephone: +43 1 58801 18839
             Telefax: +43 1 58801 918839
             Web   http://ieg.ifs.tuwien.ac.at

Abstract

MHB-F stands for Many-Headed Bridge, Free-text version. It is an intermediate representation tailored to the modelling of clinical guidelines or protocols. It is based on XML and represents the knowledge as a set of chunks, each containing a piece of information. The various aspects of each chunk are grouped into dimensions ranging from control structure to resource requirements. Breaking the original knowledge format (i.e., source document) into such chunks and marking them up along the various dimensions facilitates the later creation of more formal and more complex models, thus building a bridge between the latter and the original document.

The design of MHB-F builds on MHB, which was used in two EU projects. The further developments consist of a split into a free-text version, MHB-F, and a semantically enriched version, MHB-S. Under the new modelling scheme, the original free-text is first translated to MHB-F. Then, the MHB-F model is translated to MHB-S. This lowers the barriers for non-IT persons to join the modelling process in the first step, and it clarifies the phases of MHB modelling which were tacit in the original MHB modelling process.

Contents

1 Introduction

1.1 Formal structure

A file in MHB-F consists of a series of chunks which are grouped into chunk-groups. The top-level chunk-groups are contained in the formal root element.

Element 1 root

Child Name	Type or Reference	Occurrence	Description
zero or more of the following			
chunk	*#5, p. 5*	*exactly one*	One bit of information
chunk-group	*#2, p. 3*	*exactly one*	Used to structure the file
refer-to	*#4, p. 4*	*exactly one*	Reference to another chunk-group

Element 2 chunk-group

Child Name	Type or Reference	Occurrence	Description
title	*String*	*zero or one*	Section name or other internal categorization
group-id	*String*	*zero or one*	Unique ID used to jump to refer to this group in refer-to
document-link	*#3, p. 4*	*zero or more*	Added by editing tool. Pointing to title of section.
zero or more of the following			
chunk	*#5, p. 5*	*exactly one*	A single bit of information
chunk-group	*#2, p. 3*	*exactly one*	Another group nested in this group
refer-to	*#4, p. 4*	*exactly one*	Reference to another chunk-group or chunk or title of external document

1.2 Context of the MHB-F file and document linking

MHB-F is an intermediate representation. This means, that each MHB-F file is derived from one or more source file(s) – natural language text (free-text) describing guidelines or protocols. Also, there is at least one document or file derived from each MHB-F file in most use cases. It is called the target file here.

In contrast to similar approaches developed by other teams, MHB-F explicitly maintains the links between corresponding pieces of information in related documents, i.e., in source, MHB-F and target file. To formally represent these links, each element contains a variable number of `document-link` elements.

Element 3 document-link

Child Name	Type or Reference	Occurrence	Description
link-id	*String*	*exactly one*	ID issued by editing tool
source-filename	*String*	*zero or one*	Name of the file, to which this link points, if it points backwards, to the guideline text.
source-text	*String*	*zero or one*	Text snippet to which the link points
target-filename	*String*	*zero or one*	Name of the file, to which this link points, if it points forwards, to the MHB-F model.

To represent explicit cross-references within the same document, the `refer-to` element is provided.

Element 4 refer-to

Child Name	Type or Reference	Occurrence	Description
group-id	*String*	*exactly one*	group-id of the target chunk-group
document-link	*#3, p. 4*	*zero or more*	Added by editing tool. Pointing to title of section.

1.3   The structure of a chunk

Each chunk represents one piece of information. This may be a single sentence, or a part of a sentence. Only in rare cases, it will represent more than one sentence. The properties or aspects of a chunk are grouped along the following dimensions.

Control. Everything related to control flow, as seen in programming languages.

Data. The definition, creation and use of any values.

Time. The temporal dimension of the information described here.

Evidence. The degree of evidence or other rating associated with the statement described here.

Background. Various complex aspects pertaining to background information more than to other dimensions.

Resources. The resources required for the described activity.

Patient aspects. Impact of the described activities or information on the patient.

Information source. Formal role of the information within the source document.

The following sections describe these dimensions in details.

Each chunk has a unique ID for reference purposes. It is managed by the editing tools.

Element 5 chunk

Child Name	Type or Reference	Occurrence	Description
chunk-id	*String*	*exactly one*	used to refer to by other chunks
subject	*String*	*zero or one*	Name or description of action or event which is described here, if not clear from the context.
context	*String*	*zero or one*	Specification of context for this chunk, if not clear otherwise.
document-link	*#3, p. 4*	*zero or more*	Added by editing tool. Pointing to title of section.
zero or more of the following			
control	*#6, p. 6*	*exactly one*	Aspects in the control flow dimension
data	*#13, p. 10*	*exactly one*	Aspects in the data flow dimension
time	*#18, p. 11*	*exactly one*	Aspects in the temporal dimension
evidence	*#19, p. 12*	*exactly one*	Aspects in the evidence dimension
background	*#20, p. 13*	*exactly one*	Aspects in the background dimension
resources	*#26, p. 14*	*exactly one*	Aspects in the resource dimension
patient-aspects	*#27, p. 15*	*exactly one*	Aspects in the patient dimension
information-source	*#28, p. 15*	*exactly one*	The status of the chunk in the original guideline text

2 Control dimension

One of the most prominent aspects of a guideline is, *when* to do *what*. This is often shown (incompletely) by a drawing called clinical algorithm. *When* either relates to the condition, under which something happens, or to the temporal order of actions relative to each other. The first is generally called *decision*, while the latter can be subsumed under *ordering* of activities and/or events. A third issue is *decomposition*, i.e., the way in which larger tasks or activities are made up of smaller ones. Last not least, some tasks are performed *repeatedly* as part of other tasks.

Ordering, decomposition, and repetition imply temporal aspects, e.g., one task being performed after the other. Still, they are included here, while the temporal dimension proper deals with explicit and often quantitative descriptions of the timing of actions and effects etc.

Element 6 control

Child Name	Type or Reference	Occurrence	Description
zero or more of the following			
if-then	*#7, p. 7*	*exactly one*	A pair of condition and resulting activity plus optional modifiers
choice	*#8, p. 7*	*exactly one*	A set of related condition/activity-pairs
subtasks	*#9, p. 8*	*exactly one*	A parent task and its children
repetition	*#11, p. 9*	*exactly one*	A parent task and a repeated child task
activity	*#12, p. 9*	*exactly one*	An activity which is considered atomic, i.e., for which only a description is given

2.1 Decisions

The prototypical form of a decisions is "*if* condition *then* result". However, there are many details to consider.

- The result can be a event, an effect, or an activity but in this section we only refer to activities. An activity has a personal or personalizable actor, e.g., the physician on duty, whereas a event does not have a person as its actor or driving force, e.g., a disease. For instance, in the event of Jaundice, the actors are the Bilirubin molecules or the not yet sufficiently working liver. There is an important impact of this distinction: The events of the disease are not synchronized with the activities of the care process. Therefore, modelling both as a single hierarchy of plans is rarely feasible. Usually, only the care activities are modelled and the event of the disease itself becomes visible through the result of diagnosis steps only. This resembles the view of the physician.

  Effects are changes in some observable parameters where nothing about the underlying process is known – for instance, "X leads to decreased mortality". The rate of mortality can be observed, but everything between the (assumed) cause X and the resulting rate change is unknown. Effects are described by the dimension *background*.

  To clarify the fact that we are only referring to resulting activities here, not effects, the result simply called `activity` in the MHB-F element.

- Often a decision has many options or results depending on several inputs. E.g., age plus Bilirubin level lead to one of four treatment options. This is technically the same as a set of decisions, but it is important for human understanding to use appropriate forms for complex decisions as discussed

below. In such a case, a set of `if-then` elements is grouped by a `choice` element containing the complete set.

- In the medical domain, few statements are imperative. Most statements contain some form of vagueness or restriction. This can be qualitative such as the word "should" or it can be quantitative such as "in 70 % of all cases". All these aspects are subsumed in the dimension *evidence*. Still, they are directly embedded into the decision statement in some formal representations, which is why references to them appear below.

Element 7 if-then

Child Name	Type or Reference	Occurrence	Description
degree-of-certainty	*String*	*zero or one*	Put words like may or sometimes here
condition	*String*	*exactly one*	Precise description of condition, preferable in terms of variables defined by data definitions
modifier	*String*	*zero or one*	Any modification of the relation, such as 'should not'
activity	*String*	*exactly one*	What to do if condition fulfilled, preferably in terms of a task name described somewhere else
document-link	*#3, p. 4*	*zero or more*	Added by editing tool

Element 8 choice

Child Name	Type or Reference	Occurrence	Description
parent-activity	*String*	*zero or one*	Unique task name of parent task
selection-mode	*String*	*zero or one*	Any additional information regarding chosing between the options, such as 'only one'
if-then	*#7, p. 7*	*one or more*	A pair of condition and action plus modifiers for each

2.2 Decomposition of tasks

A parent task (or plan) P can be decomposed into children (or subplans or subtasks) A, B, and C. Unfortunately, we found that the term decomposition is not widespread nor widely understood outside computer science, at least not in the meaning that is intended here. Therefore, we opted to name this element `subtask` in MHB-F.

The ordering of the children is specified in element `ordering` inside `subtask`. Its values will be taken from the guideline text in MHB-F. However, in a later

modelling step (translating MHB-F to MHB-S), this term will be translated into an ordering available in the target representation(s).

While the form of presentation is only a formal problem, the lack of information about the precise constraints of ordering in the guideline can be an important issue. The examination of such information gaps can contribute to improved quality in guideline design even though many gaps will be intentional.

Element 9 subtasks

Child Name	Type or Reference	Occurrence	Description
parent-task	*String*	*zero or one*	Unique task name of parent task
task-description	*String*	*zero or one*	Details of the task which are not formalized
ordering	*String*	*zero or one*	Sequential or parallel etc.
degree-of-certainty	*String*	*zero or one*	Put words like may or sometimes here
document-link	*#3, p. 4*	*zero or more*	Added by editing tool
child-task	*#10, p. 8*	*one or more*	Each element represents one child task.

Element 10 child-task

Child Name	Type or Reference	Occurrence	Description
name	*String*	*exactly one*	Unique child task name
degree-of-certainty	*String*	*zero or one*	Put words like may or sometimes here
document-link	*#3, p. 4*	*zero or more*	Added by editing tool

2.3   Repetition of tasks

Often one task is performed more then once – either for a certain number of times or at certain times during another task. In any case, there is a parent task ("envelope task") which lasts during all the repetitions and a child task ("repeated task") which is performed repeatedly. The form of repetition together with all its constraints is entered in the element `repeat-specification`.

Element 11 repetition

Child Name	Type or Reference	Occurrence	Description
envelope-task	*String*	*zero or one*	Unique task name of parent task
repeated-task	*String*	*exactly one*	Unique task name of child task
repeat-specification	*String*	*zero or one*	How often and with which constraints is the child repeated?
degree-of-certainty	*String*	*zero or one*	Put words like may or sometimes here
document-link	*#3, p. 4*	*zero or more*	Added by editing tool

2.4 Atomic tasks or activities

Some activities or tasks are only described in a sentence without further detailing them. In these cases, they are generally called *atomic actions*. In MHB-F, we call them `activity`.

Element 12 activity

Child Name	Type or Reference	Occurrence	Description
name	*String*	*exactly one*	Unique task name
description	*String*	*exactly one*	Details on the activity.
document-link	*#3, p. 4*	*zero or more*	Added by editing tool

3 Data dimension

Interwoven with control flow is the description of the data processing involved in the diagnosis and treatment of the patient. While the processing of data seems of little importance in the treatment of many diseases, it is often prominently described in the diagnosis part of a guideline. As control flow describes the gathering of information, data flow describes how one piece of information is abstracted from other ones. We distinguish the following:

- The *definition* of a data item is rarely found in the guideline in explicit form. Still, it is necessary for the formal version of the guideline. It consists of a name, a type, and often a range of plausible values and a preferred unit.

- The *usage* of a data item is made explicit to varying degrees in actions described in the guideline and calculation of other values.

- The *input* of a data item is sometimes explicitly described in the description of the patient interview or diagnosis.

- *Abstraction rules* describe the calculation or abstraction of one data item based on others. It is found mostly in descriptions of diagnosis. The time

at which the calculation or abstraction is performed may be explicitly stated or not. In the first case, the statement in question has a data flow aspect and a control flow aspect at the same time. In the second case, abstraction is assumed to take place automatically whenever necessary.

In order to bring the syntax of MHB-F nearer to the user without IT background, we renamed the above aspects to the less abstract terms *description*, *used*, *produced*, and *calculation*.

Element 13 data

Child Name	Type or Reference	Occurrence	Description
zero or more of the following			
description	*#14, p. 10*	*exactly one*	Definition or description of one data entity, i.e., variable
used	*#15, p. 10*	*exactly one*	Names a variable or piece of information which is used in this chunk
produced	*#16, p. 10*	*exactly one*	Designates a place where a certain bit of information is supplied by the user
calculation	*#17, p. 11*	*exactly one*	Describes how one variable is calculated based on others in a declarative way

Element 14 description

Child Name	Type or Reference	Occurrence	Description
name	*String*	*exactly one*	Unique variable name
document-link	*#3, p. 4*	*zero or more*	Added by editing tool

Element 15 used

Child Name	Type or Reference	Occurrence	Description
name	*String*	*exactly one*	Unique variable name
degree-of-certainty	*String*	*zero or one*	Put words like may or sometimes here
document-link	*#3, p. 4*	*zero or more*	Added by editing tool

Element 16 produced

Child Name	Type or Reference	Occurrence	Description
name	*String*	*exactly one*	Unique variable name
degree-of-certainty	*String*	*zero or one*	Put words like may or sometimes here
document-link	*#3, p. 4*	*zero or more*	Added by editing tool

Element 17 calculation

Child Name	Type or Reference	Occurrence	Description
result	*String*	*exactly one*	Unique variable name
description	*String*	*zero or one*	Semantics or other information
document-link	*#3, p. 4*	*zero or more*	Added by editing tool

4 Time dimension

Both data and control flow may have temporal aspects. Their description tends to be complex if it is precise enough to satisfy the needs of formal modelling. However, the original text of guidelines only contains statements that can be a starting point for transformation by a knowledge engineer. Therefore, MHB-F simply provides a description slot to contain what is found in the guideline text. MHB-S structures this in a more formal way, in the next translation step.

Element 18 time

Child Name	Type or Reference	Occurrence	Description
description	*String*	*exactly one*	Free-text description
document-link	*#3, p. 4*	*zero or more*	Added by editing tool

5 Evidence dimension

In evidence-based guidelines, recommendations and conclusions are graded. However, different organisations use different schemata. Also, the strength of evidence is not just uni-dimensional. MHB-F avoids any restrictions in expressiveness by providing the superset of the combinations found in practice, as child elements of the `evidence` element. In detail, the different features are:

- *Grade* is generally used for evidence-based conclusions, expressing their strength or the overall quality of the evidence which lead to them.

- *Level* refers to the level of quality of specific sources of evidence, such as studies.

- *Importance* expresses the guideline authors' judgement and is more loosely linked to the level of the corresponding evidence.

In addition to the above, we provide the following slots to link to the source of information.

- *Literature-reference* names the articles where the evidence is described, as shown in the original guideline text.

- *Is-based-on-other-chunk* expresses a cross-reference within the guideline and its model.

Element 19 evidence

Child Name	Type or Reference	Occurrence	Description
is-based-on-other-chunk	*String*	*zero or one*	ID of other chunk
literature-reference	*String*	*zero or one*	Citation of article
grade	*String*	*zero or one*	Grade of evidence-based conclusion or recommendation according grading scheme of guideline
level	*String*	*zero or one*	Level of evidence in the literature according grading scheme of guideline
importance	*String*	*zero or one*	Importance of statement rated by guideline authors
document-link	*#3, p. 4*	*zero or more*	Added by editing tool

6   Background dimension

Background information describes various aspects of the topic. Some refer to a particular statement or group of statement while others are only loosely coupled to particular statements or recommendations. Also their potential to be formally encoded largely varies.

In MHB-F, the text blocks or sentences describing them are simply marked for further processing in the next step, when progressing to MHB-S.

- *Intentions* of the described actions or recommendations inform and motivate the reader about the reason of certain steps.

- *Effects* are relations between data or phenomena and other phenomena which are not seen as events or actions.

- *Relations* are similar to effects, but do not postulate that one of the two named entities is the cause of the other.

- Other *educational information* can be targeted at the physician or the patient to discuss the recommendations in detail. To some extent, it is related to evidence, as it is an alternative backing for the recommendations in the guideline. Still, there is a considerable amount of statements in the guideline, which merely contain important information without direct link to any activities.

- *Explanations* are an important subspecies of the above item. They contain information directly explaining recommendations or other (important) statements in the guideline. In an implementation of the guideline, it might be

useful to show them to the user (patient or physician) – in contrast to the education information above which is not as directly related to the actions resulting from following the guideline.

Element 20 background

Child Name	Type or Reference	Occurrence	Description
zero or more of the following			
intention	*#23, p. 13*	*exactly one*	Intention or goal of the action described
effect	*#24, p. 14*	*exactly one*	Expected effect of the action or process
relation	*#25, p. 14*	*exactly one*	Expected relation between two entities without known cause
educational	*#21, p. 13*	*exactly one*	Information included solely to educate the reader
explanation	*#22, p. 13*	*exactly one*	Explanation for related statement

Element 21 educational

Child Name	Type or Reference	Occurrence	Description
description	*String*	*exactly one*	Free-text description
degree-of-certainty	*String*	*zero or one*	Put words like may or sometimes here
document-link	*#3, p. 4*	*zero or more*	Added by editing tool

Element 22 explanation

Child Name	Type or Reference	Occurrence	Description
description	*String*	*exactly one*	Free-text description
degree-of-certainty	*String*	*zero or one*	Put words like may or sometimes here
document-link	*#3, p. 4*	*zero or more*	Added by editing tool

Element 23 intention

Child Name	Type or Reference	Occurrence	Description
description	*String*	*exactly one*	Free-text description
degree-of-certainty	*String*	*zero or one*	Put words like may or sometimes here
document-link	*#3, p. 4*	*zero or more*	Added by editing tool

Element 24 effect

Child Name	Type or Reference	Occurrence	Description
description	*String*	*exactly one*	Free-text description
degree-of-certainty	*String*	*zero or one*	Put words like may or sometimes here
document-link	*#3, p. 4*	*zero or more*	Added by editing tool

Element 25 relation

Child Name	Type or Reference	Occurrence	Description
description	*String*	*exactly one*	Free-text description
degree-of-certainty	*String*	*zero or one*	Put words like may or sometimes here
document-link	*#3, p. 4*	*zero or more*	Added by editing tool

7 Resource dimension

Each action consumes resources of various nature:

- *Personal* resources such as the working time of clinical staff.

- *Devices* such as treatment facilities.

- *Financial cost* can be given independent of qualitative or quantitative information on the above items. It includes also drug cost, etc.

Element 26 resources

Child Name	Type or Reference	Occurrence	Description
personnel-needed	*String*	*zero or one*	Care personnel required
required-devices	*String*	*zero or one*	Medical devices required
financial-cost	*String*	*zero or one*	Financial cost of activity
degree-of-certainty	*String*	*zero or one*	Put words like may or sometimes here
document-link	*#3, p. 4*	*zero or more*	Added by editing tool

8 Patent aspects dimension

While the resource dimension mostly represents the view of the care provider, there are several other general issues mentioned in a guideline which see treatment from the patient perspective.

- *Risk* connected to a certain treatment option or diagnostic action.

- *Patient discomfort* is often described as free-text and refers to undesired side effects of treatments such as chemotherapy against cancer.

- *Health prospective* can be seen as a resource which is gained by the treatment process, while it is consumed by waiting.

Element 27 patient-aspects

Child Name	Type or Reference	Occurrence	Description
risk	*String*	*zero or one*	Any concern associated with this action (considerable threats with low probability)
discomfort	*String*	*zero or one*	Pain or discomfort for the patient (foreseen but minor with respect to benefits)
health-improvement	*String*	*zero or one*	Expected benefit of the patient from this action
degree-of-certainty	*String*	*zero or one*	Put words like may or sometimes here
document-link	*#3, p. 4*	*zero or more*	Added by editing tool

9   Information source dimension

While the position of a statement in the guideline document could be considered a formal property, its status (mere explanation or featured recommendation) certainly forms an important context of its interpretation.

Adding this dimension of annotation to MHB-F supports the automatic formatting of the natural language version of the guideline. It also helps in merging models from multiple sources.

Element 28 information-source

Child Name	Type or Reference	Occurrence	Description
status	*String*	*exactly one*	Formal status of chunk within guideline, i.e., type of section where it was found
document-link	*#3, p. 4*	*zero or more*	Added by editing tool

10   Conclusion

In this technical report, we presented the syntax and semantics of MHB-F, the free-text version of MHB. It is a further development of the original MHB, offering the following advantages.

- Less structure within the various knowledge slots to simplify the modellers' learning task.

- Modifications of slot names based on user feedback to make the meaning more easily understandable.

- Use of XML Schema instead of DTD for the syntax definition.

Modelling a guideline or protocol in MHB-F is the first step on the way towards a computer-interpretable, formal model. The second step is the transformation of the MHB-F model into MHB-S. MHB-S differs from MHB-F in two ways.

- First, the various knowledge slots are extensively structured, leading to a more formal representation of the content of the information giving in free-text phrases in MHB-F.

- Secondly, the content of MHB-S slots is linked to concept in an UMLS ontology.

Details concerning MHB-S will be described in a future technical report.