21th Working Meeting of the IFToMM Permanent Commission for Standardization of Terminology
June 27 – July 2, 2005
Bardejov Spa, Slovakia

ON THE DEVELOPMENT OF AN ELECTRONIC DICTIONARY FOR IFTOMM

Antonius J. KLEIN BRETELER
Delft University of Technology, faculty 3ME/ Mechanical Engineering
Mekelweg 2, 2628CD Delft, The Netherlands
a.j.kleinbreteler@wbmt.tudelft.nl

ABSTRACT

The paper describes the working method of IFToMM commission A to develop an electronic version of the four-language dictionary. Using a design approach the demands for the users, the developers and the management are discussed. These demands include the hierarchical structure of terminology and the extendibility in future. The software structure that generates the dictionary files is introduced. The automatic link finding appeared to be a critical procedure, mainly because a referenced term can be used in a derived form. The link search algorithm as developed is able to find the majority of links automatically in all four languages. An additional missing links table, to be maintained manually, is needed to obtain a 100% match. The demands show that future developments require support of database structures and tools.

Keywords: terminology, commission A, electronic dictionary, HTML-help project, automatic link search

1. INTRODUCTION

Since 1971 IFToMM commission A works to create a four-language dictionary (English, French, German, Russian). The intention is to stimulate international publications and teaching in all Mechanism and Machine-related fields such as kinematics, dynamics, robotics, mechatronics, biomechanics, transportation etc. As stated in earlier publication [1] such a dictionary must provide:

- Classification of terms
- Explanation
- Unification of terms, mainly across different language platforms
- Synonyms and related terms

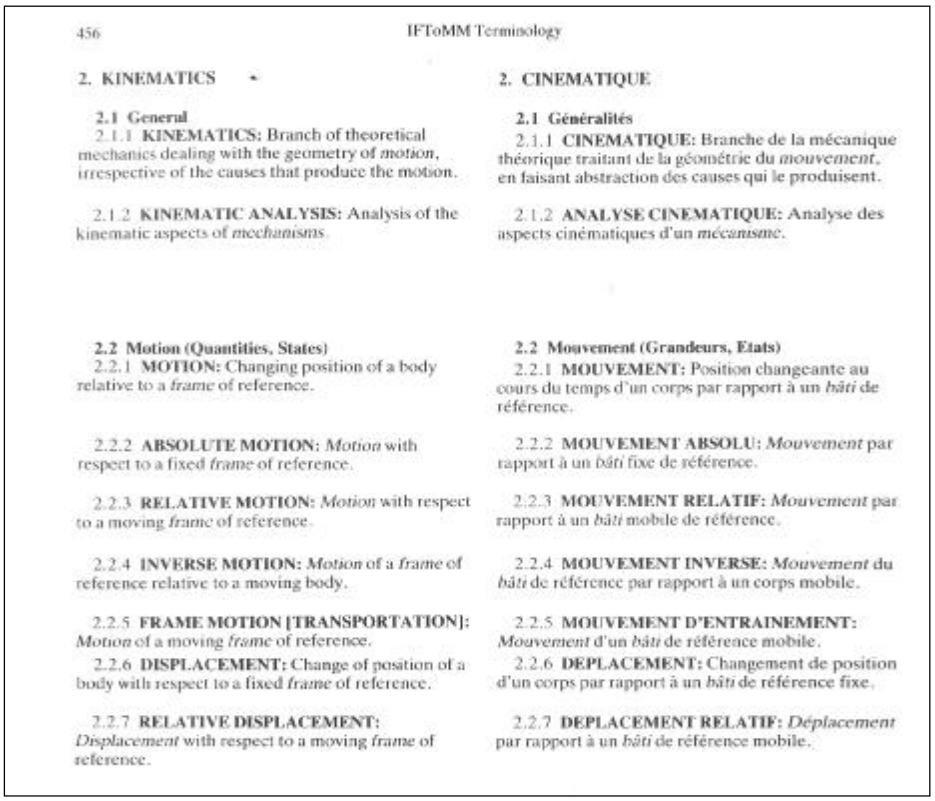
Results so far have been printed in 1991 [2] and in extended form in 2003 [3]. A sample page from 1991 is listed in figure 1. A few years ago commission A started the development of an electronic version of basically the same material, which lead in 2004 to a pre-release. A sample screen dump of the same page is presented in figure 2. The major improvement is that the referenced terms (Italics in the printed version) are now links to click on. Further the index list is available to easily search and point at a term for its explanation.
So far this electronic version should be considered as a try-out project, to see what is possible and to obtain comments. The question is then: is this the final solution? In fact the development of the electronic dictionary should be approached as any design problem, which includes here typically the following sequence of investigation:

- What should be its task/goal/function?
- What are the requirements with respect to the users, the developers, and the management?
- What are the characteristic problems to solve (analysis)?
- How can these problems be solved and how can the solutions be combined into conceptual solutions?
- Which solution is the best one, when measured with criteria reflecting its properties?
- Work out and test the chosen solution.

The intention of this paper is then to follow the design scheme when describing the development of the electronic dictionary. Since its goal has been specified already roughly, the focus will be on the user demands and the developer demands (chapters 2 and 3). Developer demands such as the adopted working method is

introduced in chapter 4 and includes the problem of extendibility of the terminology. The characteristic problem of automatic link finding is explained in chapter 5 and includes consequences for the software structure. The present achievements will be compared with the demands to detect missing functionality and discuss desired new developments

.



456 IFToMM Terminology

2. KINEMATICS

2.1 General

2.1.1 **KINEMATICS:** Branch of theoretical mechanics dealing with the geometry of *motion*, irrespective of the causes that produce the motion.

2.1.2 **KINEMATIC ANALYSIS:** Analysis of the kinematic aspects of *mechanisms*.

2.2 Motion (Quantities, States)

2.2.1 **MOTION:** Changing position of a body relative to a *frame* of reference.

2.2.2 **ABSOLUTE MOTION:** *Motion* with respect to a fixed *frame* of reference.

2.2.3 **RELATIVE MOTION:** *Motion* with respect to a moving *frame* of reference.

2.2.4 **INVERSE MOTION:** *Motion* of a *frame* of reference relative to a moving body.

2.2.5 **FRAME MOTION [TRANSPORTATION]:** *Motion* of a moving *frame* of reference.

2.2.6 **DISPLACEMENT:** Change of position of a body with respect to a fixed *frame* of reference.

2.2.7 **RELATIVE DISPLACEMENT:** *Displacement* with respect to a moving *frame* of reference.

2. CINEMATIQUE

2.1 Généralités

2.1.1 **CINEMATIQUE:** Branche de la mécanique théorique traitant de la géométrie du *mouvement*, en faisant abstraction des causes qui le produisent.

2.1.2 **ANALYSE CINEMATIQUE:** Analyse des aspects cinématiques d'un *mécanisme*.

2.2 Mouvement (Grandeurs, Etats)

2.2.1 **MOUVEMENT:** Position changeante au cours du temps d'un corps par rapport à un *bâti* de référence.

2.2.2 **MOUVEMENT ABSOLU:** *Mouvement* par rapport à un *bâti* fixe de référence.

2.2.3 **MOUVEMENT RELATIF:** *Mouvement* par rapport à un *bâti* mobile de référence.

2.2.4 **MOUVEMENT INVERSE:** *Mouvement* du *bâti* de référence par rapport à un corps mobile.

2.2.5 **MOUVEMENT D'ENTRAINEMENT:** *Mouvement* d'un *bâti* de référence mobile.

2.2.6 **DEPLACEMENT:** Changement de position d'un corps par rapport à un *bâti* de référence fixe.

2.2.7 **DEPLACEMENT RELATIF:** *Déplacement* par rapport à un *bâti* de référence mobile.

Fig. 1. Printed dictionary

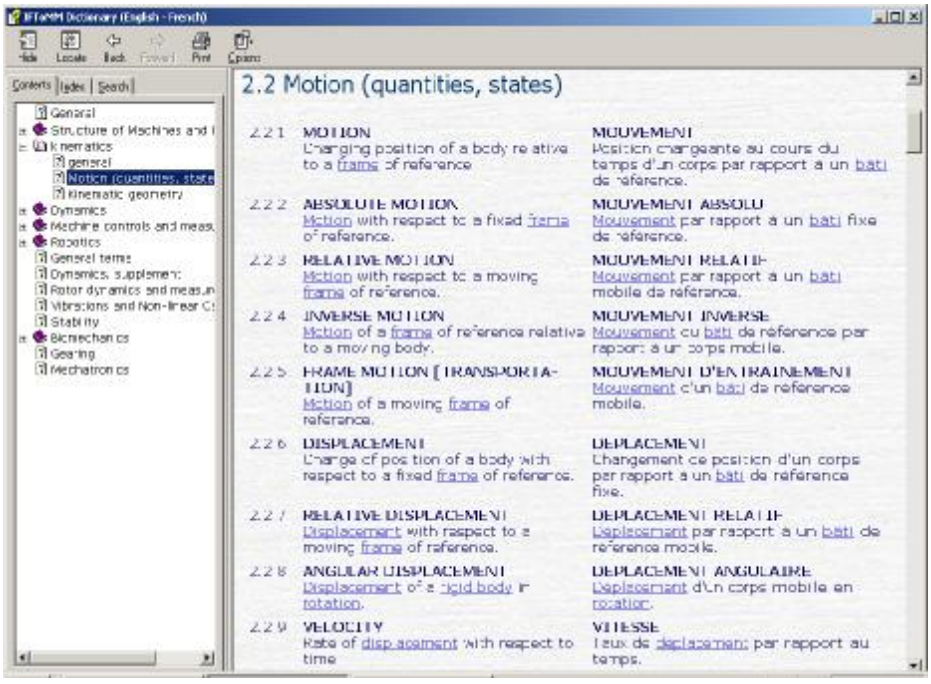


Fig. 2. Electronic dictionary

2. USER DEMANDS

In the table below the demands are collected that may be of interest for a user of the dictionary. The contents of this table are for discussion. The sign + indicates that a demand has been fulfilled already, for the present paper version or electronic version respectively. The sign indicates partial fulfilment, or fulfilment with limitation.

User demand	Paper version	Electronic version
Select set of languages		
Search for a term alphabetically	+	+
Display explanation of a term (text)	-	+
Display graphical explanation (picture)	-	-
Display reference (link)	-	+
Display reference list (all links within one term)	-	-
Display refereed list	-	-
Export of information to other channels (print)	-	+
Select content (chapter)		
Get synonyms	+	+
Advanced query options for search of a term	-	
Easy access to members of IFToMM community	-	+
Take part in discussion forum on terms	-	

Details on the fulfilment will be presented further in following chapters. It will be clear that the present electronic version is indeed a great improvement, but there are still important demands left.

3. DEVELOPER DEMANDS

The developers are not only the members of commission A, but also the contact persons and groups in the various member countries. To produce the paper version and the electronic version a certain working method has been adopted (or could be adopted). The working method should be considered together with the communication means and the working skills of the people. The demands can be listed in a table as given below. The signs + and express here the fulfilment as existing during the development.

Developer demand	Paper version	Electronic version
Effective communication means (email)	-	
Effective working method (subcommittees, standard forms)		+
Method fitting with working skills (use of text editor)		+
Intermediate products (files), compatibility		+
Processing of output (book, CD)		+

It will be clear that the introduction of the modern PC and network greatly has improved the development possibilities. But the developers certainly had to improve their own skills to take advantage of it.
The present working method will be described now in more detail, see next chapter.

4. PRESENT WORKING METHOD

From the beginning commission A decided to make certain strategic choices that allowed a straightforward working method:

- The terms to describe are divided into groups and subgroups according to a previously accepted hierarchical structure. It classifies the terms within a context of application. The hierarchical structure can be recognized in the dictionary as chapter and sub-chapter. The advantages are clear: a chapter can be connected to a subcommittee (specialists for the topics of the chapter) that takes the responsibility for the contents. The product of the subcommittees is typically one text file for each chapter and for each language, called further the **terms files**. The sequence of terms within a (sub-)chapter is guided by readability. Usually it means that the referenced terms are described earlier in the same chapter and consequently get a lower sequence number. Strict rule is that the numbering is consecutive, so each term gets a unique number.

- An alphabetic list of terms is made afterwards, one for each language. One of the members of commission A has the role of “language editor” and he produces the **index file**.
- The English definitions are discussed first, which is natural with the working language English of the commission.

The printed version of 1991 was prepared by a single editor (J. Prentiss). It contains chapters 1 - 5 and an appendix of general terms. For further development it was required that more editors can co-operate and a common platform for text editing needed to be chosen: Word97. One very good reason for this choice is that Word can handle all alphabets (Unicode, using 2 Bytes for each character). The conversion from the original material to Word was based on the transfer of plain text files (ASCII), in which some non-standard characters (like in German: ß, ä, ö, ü) were also present. The original Russian files were prepared separately and had to be edited again.

A common layout for the terms files was adopted, not only for communication reasons but also because it was recognized that the Word files must be regarded as intermediate products. So again several strategic choices had to be made, for instance:

- Use of a four-column table (term number, term definition, term explanation, remarks),
- The term definition contains also the synonyms in square brackets; see term 2.2.5 in fig. 1,
- In case of a definition using more than one word only one instance is to be used. Example: term 2.2.2 is defined as “Absolute motion” in English, where “Mouvement absolu” is used in French. It would also be possible to define the term as “Motion, absolute” in English. Such other instances are expected in the index file, not in the terms files. It is of course dependent on the preference of the language editor which of the two forms “noun, adjective” or “adjective, noun” will be used.

The term explanation text contains also links to other defined terms. Intentionally the referenced terms are placed in <brackets>, omitting the options of hypertext or Word dependent mark-up, like *Italics* in the printed edition. There are two good reasons to do this. The first reason is that links may point to different files and this would require an extra level structure in the document files. The second reason is that Word-specific mark-up will be lost when the text is exported to a plain text file. The way back should be kept open. Programming actions like providing automatic link connections should preferably be kept independent from Word-specific mark-up structures.

The improved working method served the creation of several new chapters in the dictionary. The printed version (2003) contains the new chapters 6-12, but only in English yet.

5. SET-UP OF THE ELECTRONIC VERSION

Regarding the demands for the users and the developers it was agreed that an electronic version should be about the same as the printed book, but now with additional browsing by means of links in the alphabetic index part and in the explanation part. A four-column layout, as printed in the first edition across left and right page, was preferred by the committee members. Unfortunately such a format does not match with the usual screen dimensions and an alternative idea ”two languages at a time at the screen” was accepted. The electronic dictionary is thus actually a set of bi-lingual dictionaries. Considering that each of the four languages could be regarded as “first” and “second” 16 dictionaries are needed then (including the combination of a language with itself, which appears then as a dictionary with just one column). An advantage of this approach is the possibility to develop the dictionary for additional languages. With respect to manageability and maintenance: for n languages there are n^2 dictionary files required. It is unlikely that a great number of languages can be managed, but maybe not every possible combination of two languages is required in practice.

Based on the advice and also great practical help of a professional software developer, it was decided to use the available software on the Windows developers’ platform as much as possible. It was expected that the programming job could be relatively small then. The dictionaries can for instance be created as so-called help files in a HTML-help project. The HTML-help compiler must be fed with input files having links already present. These files should then preferably be so-called XML-files, which is a standard for structured data. The programming part focuses then on the process between the Word-files and the XML-files, which includes the (automatic) link assignment. From the XML-files it is relatively easy to derive other products like PDF-files (for printing) and ordinary HTML-files (for websites).

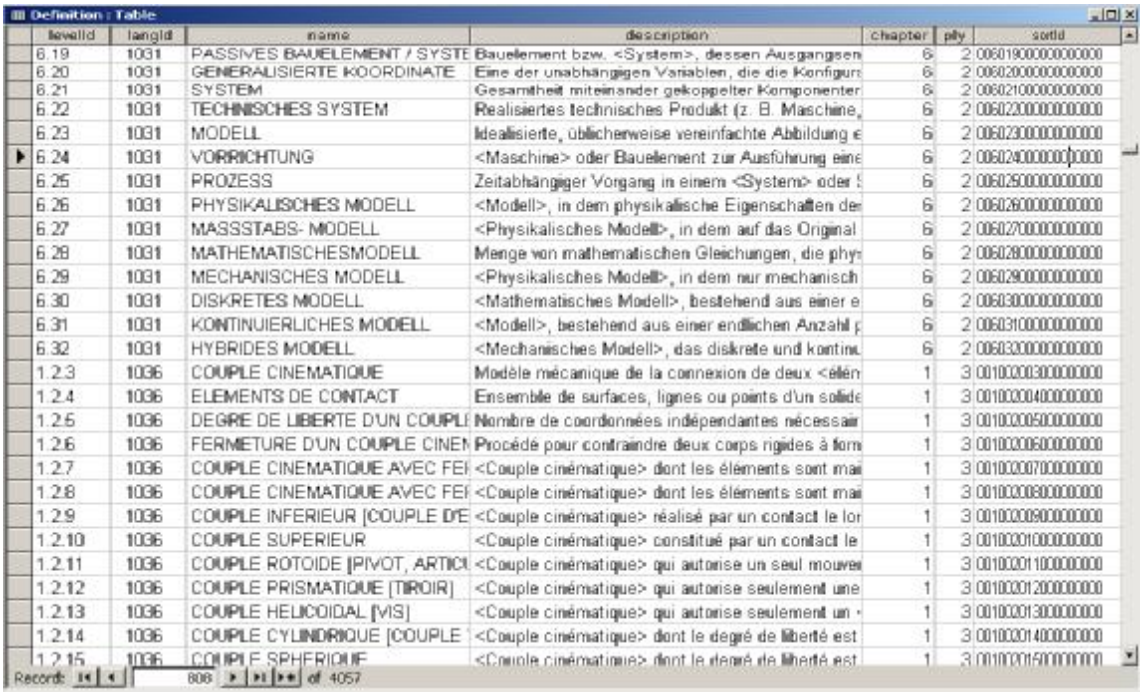
The dictionary creation software will have then a structure as depicted in fig. 5. Three stages can be recognized:

- Collection of data from the single Word files into one central database (Access). This action is required for the link search. Typically the database has two tables: one for the terms and one for the index lists. The tables can be inspected manually (sample part see fig. 3), but it is not the intention to change them manually. In case of a modification of a Word file the concerning language part of the tables must be updated and the link search must be done again.

- Creation of XML-files, in which the links have been resolved. It was chosen to make one file for each language combination and for each chapter, so there are $n^2 = 16$ files per chapter. For a sample part of an XML-file see fig. 4.
- Conversion to HTML-help files. Separately created so-called style sheets determine the actual look of the result on the screen.

Looking back when having finished the software, two problems have cost much more time than originally estimated:

– The inclusion of Russian language in the XML and HTML documents. This has been solved completely within the software.
– The automatic link resolving procedure. Several experiments with link search algorithms were performed. Results are reported in the next chapter.


Definition : Table

levelId	langId	name	description	chapter	ply	sortId
6.19	1031	PASSIVES BAUELEMENT / SYSTE	Bauelement bzw. <System>, dessen Ausgangsen	6	2	0060190000000000
6.20	1031	GENERALISIERTE KOORDINATE	Eine der unabhängigen Variablen, die die Konfigurs	6	2	0060200000000000
6.21	1031	SYSTEM	Gesamtheit miteinander gekoppelter Komponenten	6	2	0060210000000000
6.22	1031	TECHNISCHES SYSTEM	Realisiertes technisches Produkt (z. B. Maschine,	6	2	0060220000000000
6.23	1031	MODELL	Idealisierte, üblicherweise vereinfachte Abbildung e	6	2	0060230000000000
6.24	1031	VORRICHTUNG	<Maschine> oder Bauelement zur Ausführung eine	6	2	0060240000000000
6.25	1031	PROZESS	Zeitabhängiger Vorgang in einem <System> oder S	6	2	0060250000000000
6.26	1031	PHYSIKALISCHES MODELL	<Modell>, in dem physikalische Eigenschaften der	6	2	0060260000000000
6.27	1031	MASSSTABS- MODELL	<Physikalisches Modell>, in dem auf das Original	6	2	0060270000000000
6.28	1031	MATHEMATISCHESMODELL	Menge von mathematischen Gleichungen, die phys	6	2	0060280000000000
6.29	1031	MECHANISCHES MODELL	<Physikalisches Modell>, in dem nur mechanisch	6	2	0060290000000000
6.30	1031	DISKRETES MODELL	<Mathematisches Modell>, bestehend aus einer e	6	2	0060300000000000
6.31	1031	KONTINUIERLICHES MODELL	<Modell>, bestehend aus einer endlichen Anzahl p	6	2	0060310000000000
6.32	1031	HYBRIDES MODELL	<Mechanisches Modell>, das diskrete und kontinu	6	2	0060320000000000
1.2.3	1036	COUPLE CINEMATIQUE	Modèle mécanique de la connexion de deux <élén	1	3	0010020030000000
1.2.4	1036	ELEMENTS DE CONTACT	Ensemble de surfaces, lignes ou points d'un solide	1	3	0010020040000000
1.2.5	1036	DEGRE DE LIBERTE D'UN COUPLE	Nombre de coordonnées indépendantes nécessair	1	3	0010020050000000
1.2.6	1036	FERMETURE D'UN COUPLE CINEM	Procédé pour contraindre deux corps rigides à form	1	3	0010020060000000
1.2.7	1036	COUPLE CINEMATIQUE AVEC FEI	<Couple cinématique> dont les éléments sont mai	1	3	0010020070000000
1.2.8	1036	COUPLE CINEMATIQUE AVEC FEI	<Couple cinématique> dont les éléments sont mai	1	3	0010020080000000
1.2.9	1036	COUPLE INFERIEUR [COUPLE DE	<Couple cinématique> réalisé par un contact le lor	1	3	0010020090000000
1.2.10	1036	COUPLE SUPERIEUR	<Couple cinématique> constitué par un contact le	1	3	0010020100000000
1.2.11	1036	COUPLE ROTOIDE [PIVOT, ARTICU	<Couple cinématique> qui autorise un seul mouve	1	3	0010020110000000
1.2.12	1036	COUPLE PRISMATIQUE [TIROIR]	<Couple cinématique> qui autorise seulement une	1	3	0010020120000000
1.2.13	1036	COUPLE HELICOIDAL [VIS]	<Couple cinématique> qui autorise seulement un	1	3	0010020130000000
1.2.14	1036	COUPLE CYLINDRIQUE [COUPLE	<Couple cinématique> dont le degré de liberté est	1	3	0010020140000000
1.2.15	1036	COUPLE SPHERIQUE	<Couple cinématique> dont le degré de liberté est	1	3	0010020150000000

Record: 808 of 4057

Fig. 3. Part of the (internal) database, terms table

6. AUTOMATIC LINK SEARCH

The index table in the database contains all instances of the terms and is preferred as the source for the search (the terms table contains just one instance and is more complicated to investigate because of the extra information of synonyms). The index file, in Word, must therefore be free of type errors. The link text in the terms files in the <> brackets need also be error free.
The major problem of the link assignment is that the referenced term may contain a derived word, for instance a plural or a different case ending. Grammar rules are quite different in the four languages so it seems impossible to integrate all these rules in a general search algorithm. But it is expected that with some smart algorithm only a limited number of unresolved links will remain. These links can be contained in an extra “missing links” file, extra to the index file, to be created manually. This information will be used at creation of the XML-files. Further strategy is that the links in the English files can be resolved first, giving the other languages access to the English search results (term number) for extended matching trials. The search algorithm as used consists finally of the following match attempts, to be taken in the order as listed:

- Full match of all characters (without case sensitivity), search in both index table and missing links table,
- Full match except for the trailing character “s” (very effective in English, plural),
- For non-English: fair match when comparing the links of the English term (the winning term has the highest % of corresponding characters of all words to match),
- Wild search with the facilities of the Access software (sometimes successful).

The results of the search are reported to a file that must be inspected manually. In case of a missing or erroneous link the referenced text must either be modified or added to the missing links file, until all links are complete.
A discussion point is the option to anticipate for the search algorithm already during the definition of the term. When proposing a term and its description the editor should be aware of the link search procedure. It might be possible then to avoid the use of the missing links table on beforehand.

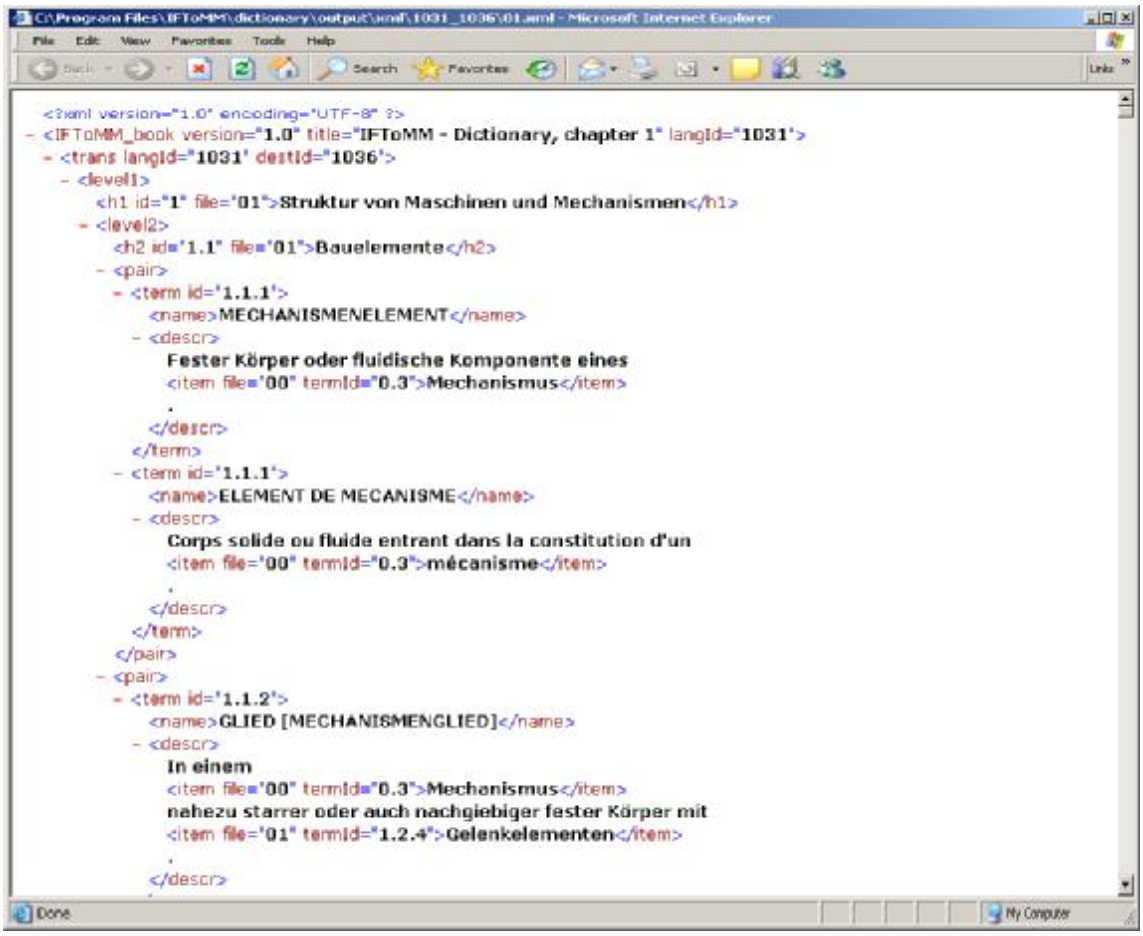


Fig. 4. Example of XML text file

7. EXTENDIBILITY

To add a new term to one of the (sub)chapters is straightforward: just add a new row at the end of the table in the terms file and give it a number one higher than the previous term. Further insert all instances of the term with the correct number in the index file.
To insert a new term somewhere in between requires renumbering of the succeeding terms. The changes in the terms table are straightforward, but the modification of the index file requires tedious attention. A possible way is using the appropriate "find and replace" functions of the text editor. The highest numbered term should be modified then first for all its instances (reversed order). The new term can be added with the "empty" number. This could be a laborious task when a lot of terms have to be renumbered.
To delete a term in a (sub)chapter requires also renumbering of the remaining terms in the chapter. Comparable to the "insert term" action the modifications in the index file must be performed one term by another, but now starting with the lowest term to renumber.
Due to the automatic link procedure extendibility of the electronic dictionary is fairly good with the present working method. Improvement may be desirable with respect to the amount of work for renumbering terms. An option is to modify the working method: create the index files not directly with the text editor, but with the dictionary program that has to be extended for this purpose. The file to create manually (as input for the dictionary software) is a simple list of all instances of terms, in numbered sequence, for each language. Such a list is much easier to modify than the index file itself. The latter method relies on the alphabetic sorting

capabilities of the database software (Access). Layout details in the index file such as concatenation marks (~ or -) can be added afterwards, but require additional programming work. The final result can be a Word file as edited manually with the present method. To adopt this method it is preferred that the syntax rules for the instances of terms are the same in all languages.

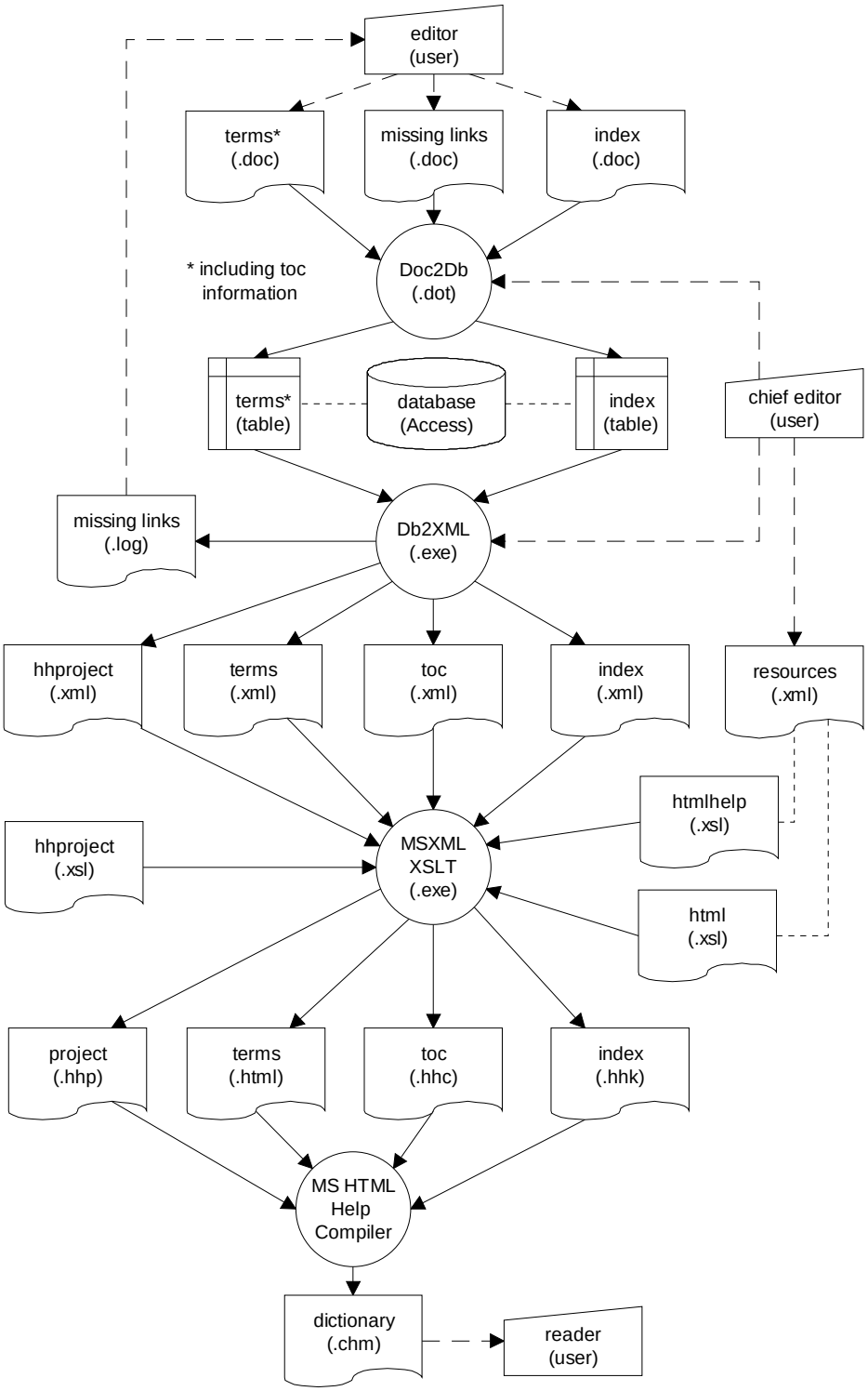


Figure 5
Structure of
the software

8. CONCLUSIONS AND FUTURE DEVELOPMENTS

The first electronic version of the IFToMM dictionary has successfully been created by commission A. The working method is based on the use of normal text editing (in Word) of tables with terms, corresponding to the printed version. These text files are input for specially developed software that creates the dictionary files and automatically adds the links to the referenced terms. The present working method makes the dictionary fairly extendable, but inserting new terms can occasionally be laborious.
From the user demands it will be clear that considerable improvements of the electronic IFToMM dictionary can be proposed. It is expected that the use of a database program can fulfil all demands as listed. This would also improve the manageability of the dictionary. Based on such an internet dictionary, it would be possible to have a discussion platform regarding the terminology for the whole IFToMM community. The role of the commission A would then be a forum of experts in terminology. A working method for such a database version must be developed, but it can be expected that the present method is a base for this.

Acknowledgement

The author wishes to thank mr. Edwin van der Klaauw, who is a senior software developer specialized in publishing software. In his free time he was willing not only to advice the software structure, but also to make the software for the multi-language dictionary. IFToMM can make free use of this software to create its electronic dictionary.

REFERENCES

[1] BOEGELSACK, G.: Twenty-five years IFToMM commission A “standardization of terminology” – History, methodology, result and future work. Mech. Mach. Theory, 33, 1998, pp. 1-5
[2] IFToMM commission A: Terminology for the theory of machines and mechanisms. Mech. Mach. Theory 26, 1991, No.5, pp. 435-539 and I-XXXIX
[3] IFToMM commission A: Terminology for the mechanism and machine science. Mech. Mach. Theory 38, 2003, No’s 7-10, pp. 507 - 1111