

SUGGESTION FOR A MORE PRODUCTIVE WORKFLOW AND INFRASTRUCTURE OF THE PERMANENT COMMISSION ON STANDARDIZATION OF TERMINOLOGY

Torsten BRIX ⁽¹⁾, Burkhard CORVES ⁽²⁾, Ulf DÖRING ⁽¹⁾

⁽¹⁾ Ilmenau University of Technology, Germany, torsten.brix@tu-ilmenau.de, ulf.doering@tu-ilmenau.de

⁽²⁾ RWTH Aachen University, Germany, corves@igm.rwth-aachen.de

ABSTRACT

The paper presents a workflow as well as an according technical infrastructure for the maintenance of the so called IFToMM dictionary. Besides the currently maintained lists of terms and definitions in four different languages the authors suggest to support some extensions of the dictionary. For users these extensions cover new languages, links to resources with occurrences of terms and definitions and explaining images. Furthermore a more formal syntax of compound words and word groups shall improve the usability of the stored information. For the maintainers of the dictionary a web-based tool would allow a more productive work by supplying revision control for items and dictionaries, access control for maintainers, change logs, discussions, statistics, reminders and last but not least automatic generation of the output formats. The web-based workflow enables all maintainers to see and retrace their results immediately.

Keywords: IFToMM Terminology, Web-based Maintenance Process, Workflow, Use cases, Motivation

1. INTRODUCTION

This paper represents the current state of our work on developing suggestions for a workflow and an according working environment for the work of the Permanent Commission on Standardization of Terminology (commission A, see [1]). At the moment there are still open questions and in some questions even the authors are of different opinions. Nevertheless this paper shall be a basis for an intensive discussion and will hopefully help to find and implement a new workflow which is capable of managing the challenges of an increasing set of technical terms, of more supported languages and of a better usability of the deliverables.

At first the authors want to give some remarks on terms which are used in this article as well as remarks on the use of computer technology. The current work of commission A results mainly in the so called IFToMM dictionary which includes *terms* and their *definitions* in the four official languages English, French, German and Russian, see e. g. [2] to [5]. Depending on the selected language each *term* may be represented by a *simple word* (e. g. *cam*), a *compound word* (e. g. *Kurvenscheibe*) or a *group of words* (e. g. *disk cam*). The *definition* of a *term* may consist of a *phrase* or one or more *sentences*.

Regarding the problems of continuously increasing lists of terms and definitions [6] demanded 2001 computer aid in maintaining the terminology as well as online publishing. According to [1] computer aid means today the use of Microsoft Word-files with a certain table format as well as the use of a computer program which transforms these Microsoft Word-files into web pages (inclusively internal links to other terms which are part of the definition). Any combination of two of the four languages is supported in the online version, where end users may find terms and definitions and also may compare the words in the two selected languages.

For several reasons explained in this article it seems to be important to move the focus from the *dictionary* itself towards to a *database* which contains all the data which are necessary to generate the dictionary in its different representations automatically (similar to the currently used computer program described above). Nevertheless from the end users point of view the terms *database* and *dictionary* may be used synonymously because he will not see the *administration information* which makes the difference between the database and the dictionary which is implicitly stored in the database.

In the following text a certain term of the dictionary and its additional information (e.g. a definition, comments etc.) is called *item*. The parts of the items (*pieces of information*) may be represented in different languages and may have different revisions. It is very important that each item can be identified independently of the provided information. In the current dictionary the identification problem is solved by a string of point separated numbers describing chapter, section and item (e. g. 1.2.15).

Besides the management of continuously increasing item lists the support of new languages (see [7]), a better use of the man power of the commission members and their supporters as well as improved output for the end users are challenges. The aim of the article is to suggest a more productive workflow of Commission A. For this purpose we:

- analyze use cases for the results of the work of Commission A (to enable a more goal oriented work),
- identify problems in the current workflow (to avoid unnecessary delays and demotivating formalities) and
- try finally to sketch an improved workflow and an according working environment.

2. USE CASES, CURRENT PROBLEMS AND FIRST CONCLUSIONS

2.1. Observation

The work of commission A will be the more productive the better it focuses its work on the needs of the (end) users. Therefore use cases of the dictionary shall be identified in this section. The use cases are extracted based on the current dictionary (like in [2] to [5]) as well as on the suggested versions of the dictionaries. Table 1 includes use cases and according features of the dictionary.

Table 1. Use cases – how dictionary content may be used by users or software applications.

1.	Search for the meaning of a certain term or group of words in some language	If the user makes such a search he usually will not be an expert. He is probably not be able to select and browse the correct chapter. The use of a search engine which supports stemming and synonyms may lead him to the matching items where he can find definitions and translations. He may use links in the definitions if he does not understand all terms used in the definition. Search should be controllable, e. g. according to the understandable languages, certain dictionary chapters or the type of information. If a narrow search in the dictionary items is not successful, then the search space may be extended to other indexed texts like documents in the suggested online library.
2.	Looking for the translation of a term or a sentence.	See 1. Here the search may be controlled in a way that only source language and destination language are shown in the hit list.
3.	Online learning of terms and definitions in one or more languages.	May be important for foreign students. Users should be able to select the needed languages (e. g. English and French), to see some explaining images and to use links to other online resources. The listed items should belong to a certain topic (e. g. a chapter in the current day dictionaries).
4.	Offline learning of terms and definitions in one or more languages.	The user may download an archive of static html pages which may be used on a computer locally or a textual document like a PDF file. Links within the exported dictionary should be supported.

5.	Translation support in online search	For web sites like DMG-Lib with multi lingual content (see [8]) it is often desirable to enlarge the set of search hits by automatically extending the query with translations of the searched terms. The more translations are usable (e. g. because they are defined in the IFToMM dictionary) the more translated terms can be added and the more content can be found by the users. Complicated or even inconsistent syntax prevents the use of some terms from the current version (see below).
6.	Translation support for online texts	Web sites like DMG-Lib may provide cross lingual translation support for the content of their online resources (e. g. when automatically translatable terms are marked in the content). Quantity and Quality of the IFToMM dictionary are for such a service also very important.
7.	External glossary in the field of (some) engineering science	Compilation of a glossary which integrates terms and according definitions from the IFToMM dictionary. For the compilation of the glossary certain IFToMM terms may be selected. Other sources may be used too. References from items in the external glossary to the original IFToMM dictionary items should be possible. If it is an online glossary then automatic updates/corrections should be possible.
8.	Link from a web-page to a certain IFToMM definition.	Some web-pages (e.g. wikipedia pages or online articles) used to reference online resources which give definitions. For those references persistent URLs are needed which support different parameters like term, language(s), version etc.
9.	Usage of provided links to other resources like more common online dictionaries	If users need more information (e. g. for translation of common vocabulary or technical terms from fields which are not supported by the IFToMM dictionary) they may use a separate part of the dictionary which includes lists with commented links to recommended web-sides. Such lists may also be used by the maintainers itself when they translate or check information in the database.
10.	Use of the (IFToMM) online library	Besides the interactive web-pages with the dictionaries in the different releases it could be useful to provide textual documents in formats like PDF. That will overcome the problems of publishing the grown dictionary in the MMT. Other related documents with clarified rights could be provided too. The integration of the content into the search index allows users to find such extra information with no extra effort.

The syntax used in the current release of the IFToMM dictionary is some times problematic for automatic utilization and for the interpretation by the users too. In such cases the resulting terms usually can also not be found in the search on [4]. Furthermore the upper case notation of terms can cause problems. Some examples for problematic definitions are:

- 1.1.17 NUT- bzw. WULST- KURVENSCHIEBE
“Nutkurvenscheibe” and “Wulstkurvenscheibe” are the resulting terms.
- 1.1.18 CYLINDRICAL [BARREL] CAM
“cylindrical cam” and “barrel cam” are the resulting terms.
- 12.21 SEHNE, ZAHNDICKEN- SEHNE, KONSTANTE (ZAHNDICKEN-)
Upper case letters prevent, that users can easily see that it must be here “konstante” and not “Konstante”.
- 6.16 AUSGANGSSIGNAL / -GRÖSSE]
“Ausgangssignal” and “Ausgangsgröße” should be the resulting terms, but possibly “Ausgangssignalgröße” too? What does the “j” at the end mean?

A main problem is, that some times corresponding terms can not be found easily in the different languages. An example is the English and the German translation of 1.1.17, which is in English one term (face cam) and in German it is represented by two terms (Nutkurvenscheibe and Wulstkurvenscheibe) which are special kinds of face cams.

2.2. Conclusions according to the public parts of the dictionary

Table 1 shows that various use cases should be taken into account. It is worth to support the different kinds of use because compared with the effort of maintaining the database the effort of writing a piece of program code which generates some kind of output is rather little. On the other hand the more possibilities exist to use the IFToMM dictionary (directly or indirectly by setting references to the dictionary) the better is the visibility of the work of commission A.

The current restriction to the four official languages will be given up step-by-step [7]. To allow the use of the knowledge and the man power of all commission members adding translated information in new languages should be quite easy. Similar the proposal of new topics and terms should be easy. That seems very important for avoiding demotivation. On the other hand the number of relevant terms should be limited. Therefore a state concept is needed (e. g. proposal, draft, rejected, released). Nevertheless proposed terms and topics shall be visible as soon as the proposer (and his supporters) want to make it available to the broad public (marked as his/their proposal). The most important demands / implications from table 1 and the problems discussed above can be summarized as follows:

- Supported views:
 - Web-pages with configurable lists of items (extended versions of the current day dictionaries [4] and [5])
 - Configurable web-page for a certain item
 - PDF files with configurable lists of items (optionally other text document formats could be supported too)
- Important configuration aspects for the different views:
 - Preferred language of the user for the static texts of the dictionary like menus, tool tips and help texts
 - Languages of the items (terms, definitions, etc.)
 - Version (e. g. a certain version or simply the last release version)
 - allowed (quality) states of the shown texts (e. g. if texts shall be shown too were the final quality check is currently not passed successfully – information from those items, which are only proposed or accepted but not released)
- More detailed demands (some necessary technical features):
 - Versioning of items and their pieces of information
 - Persistent Identifiers
 - URLs with parameters which allow certain views (see “important configuration aspects” above)
 - more easier syntax (some complex terms must be split up into explicitly given terms)
- Further features:
 - Integration of an up-to-date search engine
 - Maintained lists with links to online resources which are useful while doing translations
 - Online library with published releases of the dictionary and possibly further IFToMM relevant documents (and clarified rights of use)

In the currently used workflow (based on Microsoft Word files) the generation of the views described above (kinds of dictionaries) could be supported too e. g. by implementing according converters. But the main drawback would be, that the maintainers could not see the results of their work immediately (the own results as well as the work of their colleagues). A web-based workflow would overcome this drawback, because the resulting dictionaries can be generated on the fly. Furthermore it is for instance easier to:

- get feedback in case of inconsistencies,
- discuss about certain terms (and optionally store the discussion for internal use),
- send tasks/reminders or lists with problematic terms to colleagues
- and to generate statistics.

3. DETAILS OF THE PROPOSED DATA MODEL AND WORKFLOW

To support more use cases the data model should be extended. Therefore more kinds of information must be supported and the syntax of terms must become simpler (changing from implicitly encoded information to explicitly given information). Further aspects handled in this section are the life cycles of items (stages, revisions, releases), the different roles of maintainers and finally possibilities for the technical realization of the workflow as well as open questions.

3.1. Suggested kinds of information stored for an item

Each item may be seen as a list of pieces of information and some administrative data (like the unique ID and the flags storing properties like *internal*, *accepted*, *obsolete*).

Supported kinds of information should be:

- term,
- definition,
- general comment on the item (to be shown to the end users if reasonable),
- image (uploaded file on the server in a browser compatible format like PNG),
- link (to an internal or an external resource with occurrences of terms, definitions, further information),
- internal discussion about an item (should be made in English).

Further possible kinds of information:

- the “classical” item number as shown in the present day IFToMM dictionaries,
- examples (phrases, sentences where the term is used),
- comparison, distinction, delimitation of items (will explicitly address the differences between items, should be shown in all items referenced by the comparison),
- relationships between terms like *A is a B* and *C is part of D*.

Each piece of information includes:

- the ID of the information within the item (to allow the synchronization between the different translations),
- the text,
- the language in which the text is given,
- the release number and the revision number within this release,
- flags (e. g. *internal*, *accepted*, *obsolete*),
- (optionally) a reference to an information it is attached to (e. g. a definition for a certain term),
- (optionally) an internal discussion.

3.2. Suggestion for a more easier syntax

To improve readability for human readers as well as applications following suggestions are made:

- Split terms to get each alternative explicitly. Only the pure terms should be allowed.
- Each split term may have its own definition.
- Terms with no direct translation are circumscribed or suggested in a plausible way.
- Preferred terms could/should be marked with a flag to distinguish them from circumscriptions. An other evidence for a circumscription is that usually occurrences in articles, textbooks or specialist books will not be found.
- Example for solving problems in 1.1.17:
 - split the term into 3 terms (here called 1.1.17.a to 1.1.17.c)
 - 1.1.17.a: 1.term face cam {!} Stirnkurvenscheibe
 2.term grooved or ribbed cam Nut- oder Wulstkurvenscheibe {!}
 3.relation is-a 1.1.16
 4.comment 1. is not usual in German. 1. ist in Deutsch nicht gebräuchlich.
 - 1.1.17.b: 1.term grooved cam Nutkurvenscheibe
 2.relation is-a 1.1.17.a
 - 1.1.17.c: 1.term ribbed cam Wulstkurvenscheibe
 2.relation is-a 1.1.17.a
- Synonymous terms should always be stored in the same item (see 1.1.17.a above), possibly with different definitions.

Note that keeping the pieces of information rather small eases the synchronization of their content/meaning in the different languages.

3.3. Possible life cycles of a chapter, an item and information

Proposing a new chapter is related with proposing a list of relevant items. The focus of this section lies on the live cycle of a single item. The suggested concepts for item handling (stages, visibility etc.) may be in general assigned to chapters too.

The current development state of standards is usually reflected by stages. [9] gives an example for the stages during the development of ISO standards. From the stages described in [9] (proposal, preparatory, committee, enquiry, approval, publication) surely not all are needed for the work of commission A. For the live cycle of an item we propose stages as shown in figure 1.

Commission A may strive for perfectness, but obviously the work will never be perfect in quantity and quality. The challenge concerning quantity is to find the balance between accepting and rejecting proposed items. Each accepted item holds the danger of getting a too large database which becomes more and more unmaintainable. On the other hand each rejection holds the danger of demotivating the proposer and his supporters. A compromise seems to be the possibility to hold the item in the *proposed* stage as long as not all active members (inclusive the proposer and the former supporters) reject it. Items in proposed stage may be handled internally like items in the *draft* stage. That means that the maintainers can also add, delete and change information for that item in the different languages. Thus unnecessary delays can be avoided. Furthermore information of the proposed item may also be shown (tagged as proposal) to end users as long as the information is not marked as internal. The main difference between the proposal stage and the draft stage is that the maintainers (and in case the information is not only internal the end users as well) can see that the proposed term is accepted by the commission. This way the term becomes more relevant, e. g. for maintainers translating the terms into other languages or for end users browsing the item lists which show the work in process.

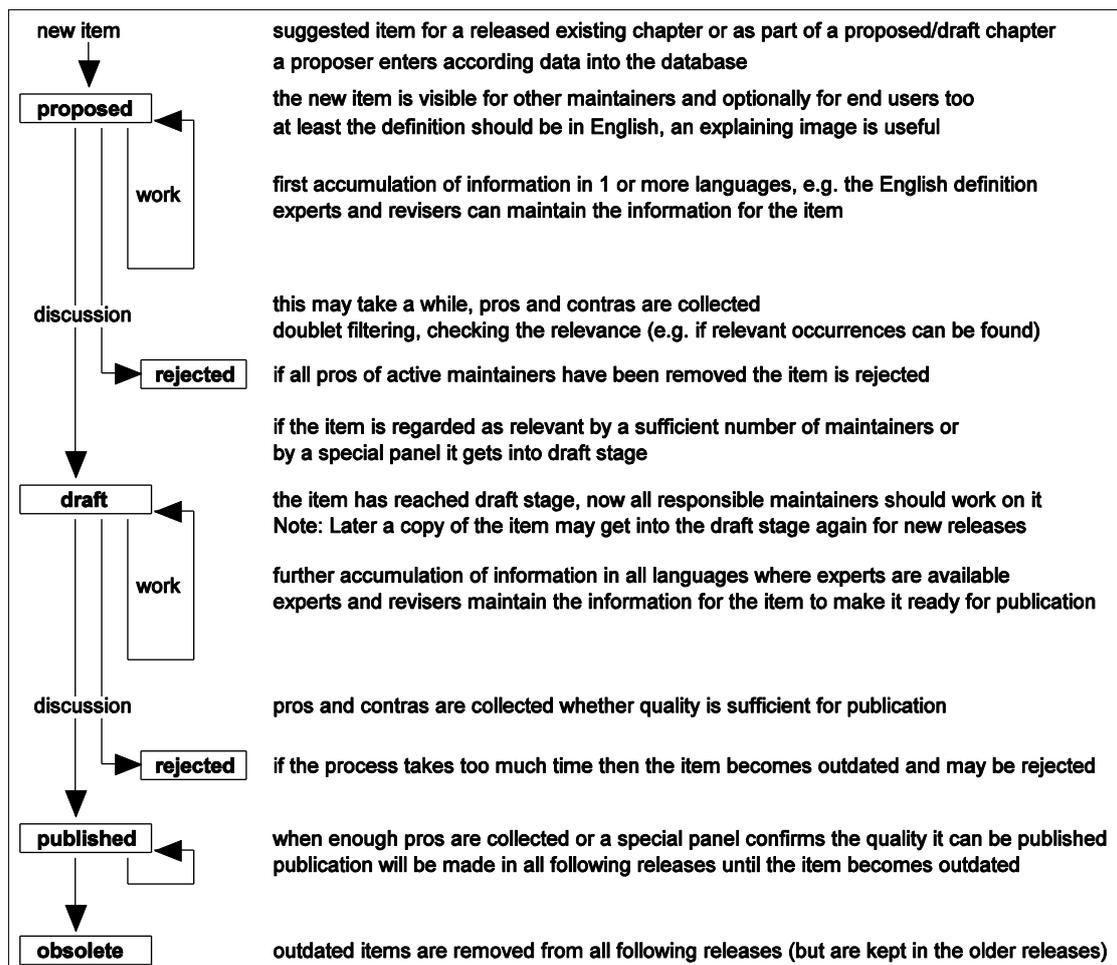


Fig. 1. Stages of an item in its live cycle.

The quality of items is reflected in two ways. One is the current stage of the item and the other one is the presence or absence of certain flags for each information. A flag is a property which may be true or false:

- *internal*: Because the internal-flag is very important for the external visibility of the information the maintainers need a strong visual reminder, e. g. two significantly different background colors for internal and non-internal information.
- *accepted*: Changes made by some expert lead to a new revision of the information. Each change shall be checked by a reviser. The reviser may accept or reject the change. The last accepted revision serves as current revision.

Both flags may be combined. The end users will only be able to see revisions which are non-internal and accepted. The distinction between quality check (accepted-flag) and visibility (internal-flag) could be seen as not necessary. Then end users might see each accepted change immediately. But especially for the earlier stages (proposed and draft) the internal-flag and the according visibility control seems to make sense. An easier way of controlling the visibility could be to place the internal-flag only on item level. Then all the accepted information for an item would be shown or it would not be shown.

If the commission A or a sub-commission decides to publish a new release then the according items are in the stage *published*. Note that information in *published* items must not be switched from non-internal to internal, because published information should never disappear. Instead an error ticket should be appended and fixed by the responsible maintainer later. To remove items in the next release they could be marked as *obsolete*.

3.4. The release problem

In general each change of a piece of information for an item could become visible to the end users at the same time it is accepted. This would for instance speed up error handling and allow the maintainers to present their work immediately to the end users. But from the classical point of view a new release (in the sense of making changes visible to end users) implies a significant change. The question is in which context or at which granularity that change should be significant. Possible granularities are the dictionary, a chapter, a single item or a piece of information. Furthermore a new release could be made regarding certain languages or for certain kinds of information.

A good set of rules for releases could be:

- Major releases of the dictionary as well as bug fixes are supported in parallel.
- Besides the last published major release all older major releases are kept on the server too, but explicitly marked as old version.
- The next major release is the union of all non obsolete items in the published stage as well as all valid items in the draft stage.
- An item is valid if it has at least an English term and an English definition.
- New items are only published in the next major release. Until this release end users can see the new items as draft (if it is not only internally visible).
- Items marked as obsolete are removed in the next major release. They are kept in the database as part of the old releases.
- In published releases usually only error corrections should be made. Changes caused by new spelling reforms will only affect the next release.
- If an item is obsolete or a piece of information is not part of the current release then further information may be added to the item, even it is in an already published release.

Adding new information to already published releases may be seen as critical (“Like printed books old releases should never been changed.”). But if the information seems to be very important for that release (important remark, etc.) or if it belongs to an item or certain information which will not be part of the next release it makes sense:

- In general the collection of information for historical items (terms, definitions etc.) which are only part of older releases becomes possible. For example comments to older (now obsolete) definitions could be given or references to newly found occurrences in online documents could be stored.
- Sometimes error corrections are not only changing some letters in an existing piece of information. It also may be necessary to give some (additional) explanation / comment to the users to avoid incorrect use of correct (or corrected) information.

3.5. Roles of maintainers

As shown above it is important to take the motivation of the maintainers into account. In the workflow this is considered by allowing the publication of proposals and drafts.

For a productive workflow it is necessary to identify the main tasks and give support where it is reasonable.

Different roles of maintainers can be identified:

- proposer (of a new item)
- supporter (for a proposed item)
- supervisor (allowed to change stages, if the necessary criteria are fulfilled)
- expert (for a certain topic and language, translates and maintains the translations, handles error tickets)
- reviser (checks at least formal criteria, should also be able estimate the quality of the translation)
- administrator (assigns roles to persons)

For the quality of information in a certain language at least two persons should be responsible. This may be problematic for languages and topics where too few IFToMM members are present. If the reviser has no or only insufficient language skills he should at least check formal criteria like the character set (e. g. UTF-8).

One maintainer may play different roles.

3.6. Technical aspects of the Workflow

Like the public online portal of the dictionary the database will be used via a web-browser. There are two main aspects:

- allow effective entering/changing of the data for the maintainers in the different roles and
- support the controlling of the workflow (avoid unnecessary delays by messaging, ensure quality, etc.).

Messaging can be done by applying tickets (which specify tasks) to certain maintainers, e. g. changes in the English definition could be reported automatically to all experts for this item in the other languages. If later an expert adapted the definition the responsible reviser could be notified automatically. An other messaging concept is the subscription. Users may subscribe to a certain kind of tasks and can see/process the according list of open tasks. Independently on the used concepts it should be possible that maintainers can activate an automatic reminder, which generates an email when certain criteria (e. g. min. number of a kind of tasks in a time interval) is fulfilled. Nevertheless reminding must be controllable else it could become a kind of spam. For different roles statistics are an important tool to check the degree of completeness of the translation for a set of items/information into a certain language. Having an overview problems may be detected and addressed easier. Besides a training and support by email, telephone, ... it is useful to provide HTML-pages with helping/supporting information on contact information, responsibilities, organigram, rules, representative examples for good and bad data etc.

Information about good common online dictionaries is collected on the portal. Of course maintainers may use these dictionaries in their work too. Nevertheless in the future support will be given more and more by spell-checkers which are integrated in the web-browser or in the operating system. Further issues are:

- Some times tasks are delegated from an official maintainer to some staff member. The Maintainer is still responsible and therefore his name is attached to the changes. If a staff member produces high quality information it should get a login and become an official maintainer too!
- Items and according information in proposals could be accepted by default. Then only the internal-flag would control the visibility.

3.7. Further open questions

- How long would be an appropriate release cycle (criteria: Significance of the changes, preserve the actuality, amount of work, necessary/expectable delays in the workflow etc.)?
- Does it really make sense to show information which is not finally released? Can that motivate maintainers? Can it help end users? Can it generate external (end user) feedback for maintainers?
- How many maintainers usually work on a certain chapter / language? How could external experts be involved into the processes?
- Would you expect technical problems for the work via Internet? How important will be fallback strategies? Which kind would you suggest?
- What kind browser are you using? Does it support JavaScript?

- Should end users be able to give hints, e. g. missing aspects in definitions (example: 1.1.16 addresses translation in the German definition but not in the English one)?
- Are context sensitive definitions a hint for items which should be split (see e. g. 12.25 and 12.26)?
- Which steps should be made first? May be that the kinds of information given to items should be restricted to terms and definitions only. New kinds of information could be added later.

CONCLUSIONS

Based on the suggestions made in this article and based on a prototype of the web-based database the authors hope to start a constructive and stimulating discussion with the other members of commission A.

REFERENCES

- [1] Official web-pages of the IFToMM PC on Standardization of Terminology. Hosted on servers of the Queen's University, Kingston, Canada, last visited on 20 April 2010, http://130.15.85.212/terminology/PC_StandardsTerminology.htm
- [2] IFToMM Commission A: Terminology for the Theory of Machines and Mechanisms. Mechanism and Machine Theory, Vol. 26, No. 5, pp. 435-539, 1991.
- [3] IFToMM Commission A: Terminology for the Mechanism and Machine Science. Mechanism and Machine Theory, Vol. 38, Nos. 7-10, pp. 598-1111, 2003.
- [4] IFToMM PC on Standardization of Terminology: IFToMM dictionaries online. Web-pages of the Delft University of Technology, Netherlands, last visited on 20 April 2010, <http://www.iftomm.3me.tudelft.nl>
- [5] IFToMM PC on Standardization of Terminology: IFToMM dictionaries online. Web-pages of the IFToMM hosted on servers of the Queen's University, Kingston, Canada, last visited on 20 April 2010, <http://130.15.85.212/terminology/TerminologyWeb/index.html>
- [6] Ionescu, T.: IFToMM Permanent Commission on "Standardization of Terminology" – THIRTY BEAUTIFUL YEARS. Newsletter of Commission A 2001, webpage, last visited on 20 April 2010, <http://130.15.85.212/newsletter/Newsletter01/StandardsC.htm>
- [7] Klein Breteler, A. J.: Newsletter entry from PC Standards and Terminology, December 2009, last visited on 20 April 2010, http://130.15.85.212/newsletter/Newsletter09/Report_PC_Terminology_09.pdf
- [8] Web-pages of Digital Mechanism and Gear Library. Last visited on 20 April 2010, <http://ww.dmg-lib.org>
- [9] ISO: Stages of the development of International Standards. Last visited on 20 April 2010, http://www.iso.org/iso/standards_development/processes_and_procedures/stages_description.htm