Multilingual illustrated μ-thesaurus of "mechanism" indexing terms

E.-C. Lovasz[*]
Politehnica University of Timişoara
Timişoara, Romania

D. Perju[·]
Politehnica University of Timişoara
Timişoara, Romania

B. Corves[§]
RWTH University Aachen
Aachen, Germany

T.Brix[**]
Technische Universität Ilmenau
Ilmenau, Germany

K.-H. Modler[··]
Technische Universität Dresden
Dresden, Germany

A. Lovasz[§§]
Politehnica University of Timişoara
Timişoara, Romania

Abstract — *A multilingual illustrated μ-thesaurus of "mechanism" indexing terms ensures a coordinated and uniform indexing of technical information in the field of mechanisms, such that information can be accessed in the shortest possible time and language barrier be overcome. This paper proposes a complete database of illustrated science-specific mechanism indexing terms, which will be integrated into a digital mechanism library. The thesaurus uses the mechanisms science terminologies, which are included in the IFToMM Dictionary.*

Keywords: mechanism science, coordinated indexing, multilingual thesaurus.

I Introduction

Due to the information explosion that took place in the digital age an increasingly controlled indexing of information of any type is required regardless of the medium on which it is stored. In the traditional library bookstore indexing was performed through classification, indexing term is a n umerical code like: Universal Decimal Classification, the Dewey Classification, Library of Congress Classification, etc.

Information indexed using classification indices can be found in the traditional catalogs. The era of computerized indexing requires the use of indexing terms such as simple or complex phrases, different indexing languages.

The most known indexing systems on international topics are: LCSH - Library of Congress Subject Headings [1], CSH - Canadian Subject Headings [2], RVM - Repertoire de vedettes-matiere [3]; RAMEAU – encyclo-pedic authority file maintained by the National Library of France [4]; MeSH - Medical Subject Headings [5], etc.

II Indexing methodologies

Indexing effectiveness on subjects depends very much on quality of indexing language. This language should be very familiar to library users. Indexing is a complex process in which the following elements are present: indexator, indexing language, the term(s) indexing information to be indexed. It is defined as representation, using natural language or documentary data resulting from analysis of the contents of a document or part of a document to its retrieval or analysis type of a document whose content is expressed by descriptors or key words [6].

Indexers are specialized in indexing individual documents, which can be an individual or a machine/ computer. If an individual is indexer, it is called intellectual indexing, and if the indexing is done by computer, it is called automatic indexation. The intellectual and the automated indexing may be free or controlled. Indexing is a method of indexing freely by accepting the terms of the documents, or by setting the index without using a controlled vocabulary [6]. Indexing is performed in three steps which in practice tend to overlap: examining the document and define the topic, identifying and selecting key concepts of the subject, expressing concepts in terms of indexing [6].

An index term may appear both as a term derived from natural language, preferably simple or compound noun or as belonging to a p articular classification code bibliographic classifications. If the index term has the form of a classification code, it is a classification. After examining the document, indexator should identify concepts, key elements of content description, according to an identification scheme, based on criteria identified as important in the area of indexing.

The choice depends on concepts from printed alphabetical index to an automatic bibliographic database. The number of indexing terms should not be arbitrarily limited, but determined by the amount of information contained in the document related to information requirements of potential users [6]. Once the concepts have been identified and selected by the indexator for the document that will be indexed, she/he will proceed to the choice of indexing terms. To convert the concepts in terms of indexing the following rules must be observed:

- the concepts that are already in the indexing language must be expressed in terms of preference;
- terms which are new concepts must be checked in terms of content and form with reference tools

[*] erwin.lovasz@mec.upt.ro
[·] dan.perju@mec.upt.ro
[§] corves@igm.rwth-aachen.de
[**] torsten.brix@tu-ilmenau.de
[··] karl-heinz.modler@tu-dresden.de
[§§] agneta.lovasz@mec.upt.ro

(dictionaries and specialized encyclopedias, thesauri of normalized terms, classification tables) which can be consulted by specialists in the field, especially when they have knowledge of indexing.

The indexer as an individual must meet the following demands: determining the document's subject (topic), choosing words which taken together characterize the subject as well as indicating the relationships between concepts expressed by these terms.

The indexing language is a controlled set of terms selected from natural language to represent in condensed form the content of documents. Indexing language can be:

- indexing language with hierarchical structure,
- indexing language with combinatorial structure,
- natural language indexing.

Languages that have a hierarchical structure (tree) are called classification. Languages having a combinatorial structure coordinates are nothing more than coordinated indexation. These indexing languages can be categorized into: language of descriptive type and language of vocabulary type. Descriptive language is a standardized language used to describe the contents of the documents, whose descriptors are taken from natural language. Alphabetically ordered list of terms represents the controlled vocabulary, which changes into thesaurus in a higher stage of processing.

III Structure of multilingual illustrated μ-thesaurus

Thesaurus is a controlled vocabulary and flexible of terms, bound together by semantic and generic relations, referring to a certain field of knowledge [6]. Thesauri can be general or specialized on different domains. Specialized thesauri on domains can be called μ-thesauri, and these may be included in General Thesauri called macro-thesaurus.

If the language used in indexing is of thesaurus, the term of indexing is considered descriptor. The indexing term represents the notion of noun form, with or without determinants. Indexing terms may be simple or compound. For compound terms the indexing morphological term can be decomposed into separate components, expressed by nouns that can serve independently as indexing terms. If language indexing is a thesaurus, specific terms will be chosen to represent a particular concept. For controlled indexing the following abbreviations are used for indexing terms based on a thesaurus:

DES – Descriptor: the term or the group of terms, which are used for controlled intellectual indexing with indexed language;

NE – Explanatory note: the note associated with a term to indicate its meaning within an indexing language;

VG – Generic Heading: the term that follows this symbol expresses the general concept which subordinates the given term;

UP – Non-descriptor: the term used for this symbol is non-descriptive (synonym of the word or symbol above, regarded as descriptive term);

TG - Generic Term: the term that follows this symbol expresses a wider concept than the given term and is placed in terms of hierarchical ascent immediately afterwards;

TA – Associated Term: is a term which designates several common properties of more indexing terms [6].

Multilingual μ-thesaurus presumes the followings: selecting the indexing terms, the structure of indexing language, establishing the descriptors and the translation of the thesaurus. To establish the multilingual illustrated μ-thesaurus were considered the following elements: DES, NE, VG, TG, TA, UP (see above), LE - translation of the term in English; LG. - translation of the term in German; LF. - translation of the term in French and IM – the image of the descriptor.

The elements LE, LG, LF an IM is not a standard indexing abbreviation, but in this multilingual illustrated μ-thesaurus it is an important element to multilingual and visual communication.

IV Multilingual illustrated μ-thesaurus database

Through selecting the indexing terms, according to the IFToMM Dictionary [7, 8] the structure of multilingual illustrated μ-thesaurus database was established. The structure of mechanisms μ-thesaurus contains nine elements with corresponding fields. The name of the fields is the same with the abbreviations of the elements. To this was added also the drawing (picture), if for each indexing term a descriptor exists. In addition to the ten fields established, the following fields was defined: **id** - single identifier (numeric field); **file_path** field for design if necessary; **file_DIGI_DB** - field for connection to digitized information, **file_STAND_DB** - field stands for the link to the scanned image, **file_A500_DB** - field for connection to the database of the Central Library of UPT (Universitatea "Politehnica" din Timişoara).

Fig. 1. Illustrated μ-thesaurus database structure

The application used is XAMPP, where X (it reffers to the large number of operating systems that offers support), A (Apache), M (MySQL), P (PHP) and P (Perl). Microthesaurus database structure is given in Figure 1.

Fig. 2 Fields in μ-thesaurus database

All fields with necessary information about each indexing term of the illustrated multilingual μ-thesaurus were made according to Figure 2.

The field DES - Descriptor contains the specific terms or a group of terms, which are indexed now in Romanian language. These terms are for example: "cama"= "cam", "acceleraţie unghiulară"="angular acceleration", "mecanism cicloidal"="cycloidal mechanism", forţă axială"="axial force", etc. All these terms are used as specific mechanisms science terminologies and they are included in the IFToMM Dictionary [7, 8].

The field VG - Generic Headings contains in the mechanisms μ-thesaurus three general concepts: Machine/Apparatus, Mechanism and Technique, which subordinate all the specific mechanism terms.

As hierarchical ascent terms of the descriptor, TG – Generic Term, the specific chapters of mechanisms science: structure, kinematics, synthesis, kinetostatics and dynamics terms were used in the database field. This field can have one or more generic descriptors. For example for the descriptor "cam mechanism" can introduce all the defined generic terms, but for the descriptor "degree of freedom" only the generic term "structure" can be found.

The associated term TA defines a common property of more terms, for example the associated term "element"

must find the descriptors "cam", "rocker", "spur gear", etc.

In the case that the descriptor has a synonym – non-descriptor, this will be defined in the field UP.

The fields LE, LG and LF contain the corresponding terms of the descriptor in the English, German and French language.

For each indexing term of the μ-thesaurus a definition or some explanations are given, which are introduced in the field NE - Explanatory Note.

If the mechanism term has specific illustrated or drawing symbol, this will be loaded as picture as a gif-file in the field "file_path".

After completing the information about each indexing term, this term will be validated.

The interface of the multilingual illustrated μ-thesaurus database is shown in the Figure 3.

Fig. 3 Interface of the μ-thesaurus database

In the multilingual illustrated μ-thesaurus database are introduced 385 mechanism science indexing terms.

V Using examples of illustrated μ-thesaurus database

Example1

If, for example, the indexing term "cam" is to be found, than this term is typed in the search field, and the result is shown in the Figure 4. The entry fields for the advanced search engine are not activated.

The search results reveal more indexing term or group of terms (descriptors), which contain the word "cam". In this example 16 terms are found: "arbore cu cama" = "camshaft", "camă" = "cam", "camă sferică" = "spherical cam", "camă cilindrică" = "cylindrical cam", "camă de translaţie" = "translation cam", etc.
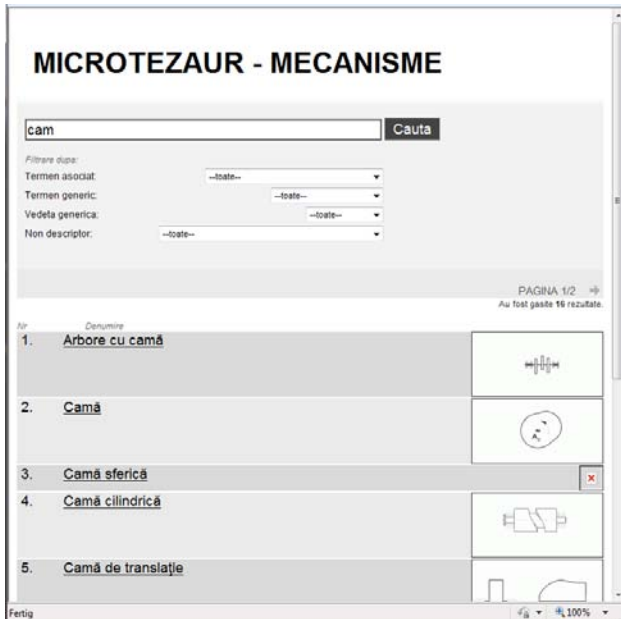
3

Fig. 4 Interface of the μ-thesaurus database

Each resulting descriptor is accompanied on the right side with his specific illustrations or drawing symbol (see Figure 4). The user can select the desired descriptor among the resultant descriptors, which contained the word "cam". For example the chosen descriptor is "camă" = "cam". The result is a new window (see Figure 5), which is dedicated to the chosen descriptor, in the example "camă" = "cam".
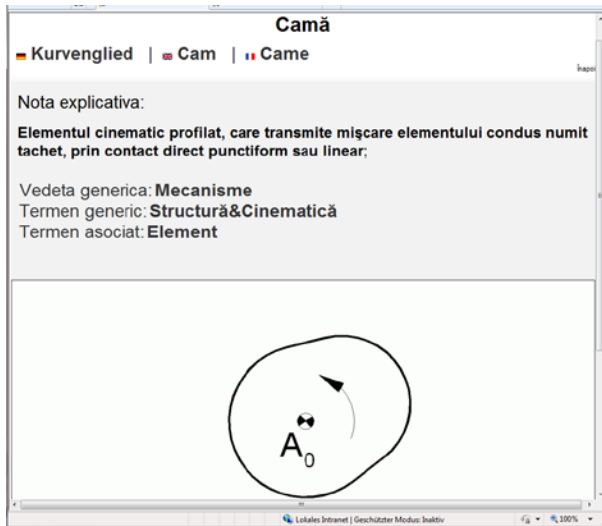

Fig. 5 Specific window of the descriptor "camă" = "cam"

The window of the descriptor "camă" displays all the information about the specific indexed term "camă" from the database. The fulfilled fields introduced in Romanian

language of the database in this example of the indexing term "camă" are the following:

DES: Camă

NE: "Elementul cinematic profilat, care transmite mişcare elementului condus numit tachet, prin contact direct punctiform sau linear"

VG: Mecanisme

TG: Structură

TA: Element

UP:

L.E.: Cam

L.G.: Kurvenglied

L.F.: Came

File_Path: C:/thesaur/cama.gif

Example 2

To illustrate the use of the advanced search in the μ-thesaurus database, it will set only in the search field of associated term TA the word "element" (see Figure 6). The search results are 36 indexing terms. The first five shown are: "arbore cardanic" = "cardan shaft", "arbore cu camă" = "camshaft", "arbore
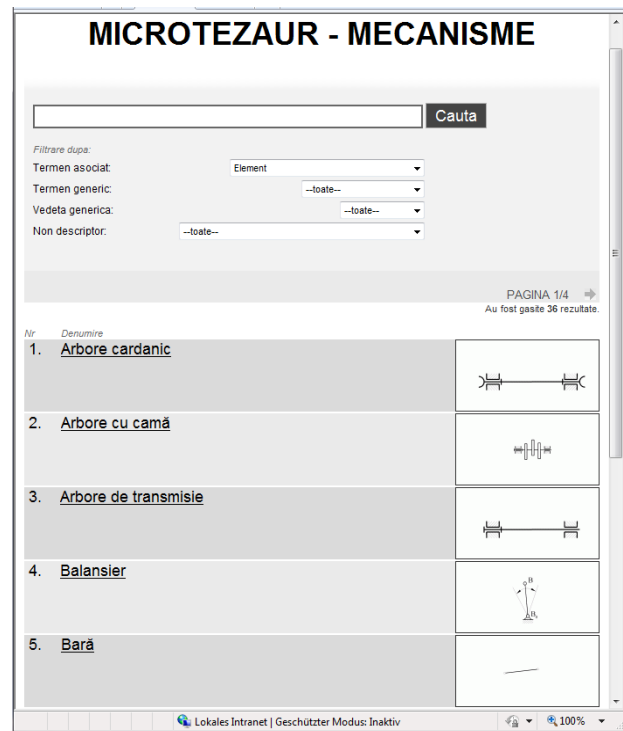

Fig. 6 Interface of the μ-thesaurus database for the advanced search on the associated term "element" = "element"

de transmisie" = "drive shaft", "balansier" = "rocker", "bară" = "bar", etc.

The user can now select one of the descriptors and for each of them there is specific window, as shown in the Figure 7 and Figure 8 for the chosen descriptors "arbore cu camă" = "camshaft" and "balansier" = "rocker" respectively.
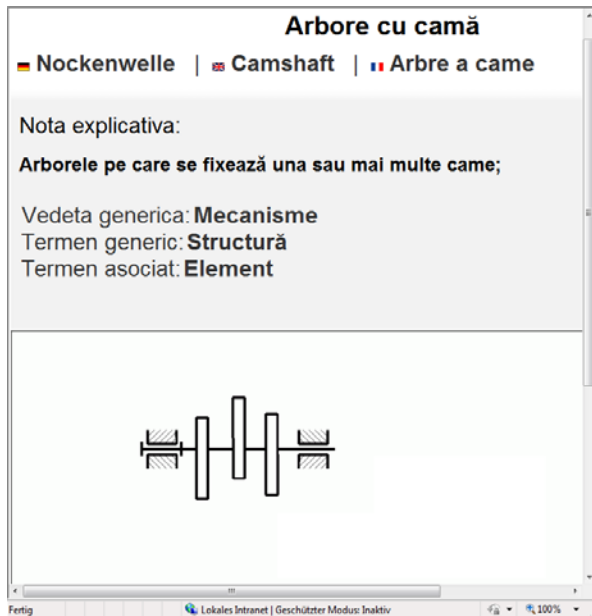

Fig. 7 Specific window of the descriptor "arbore cu camă" – "camshaft"
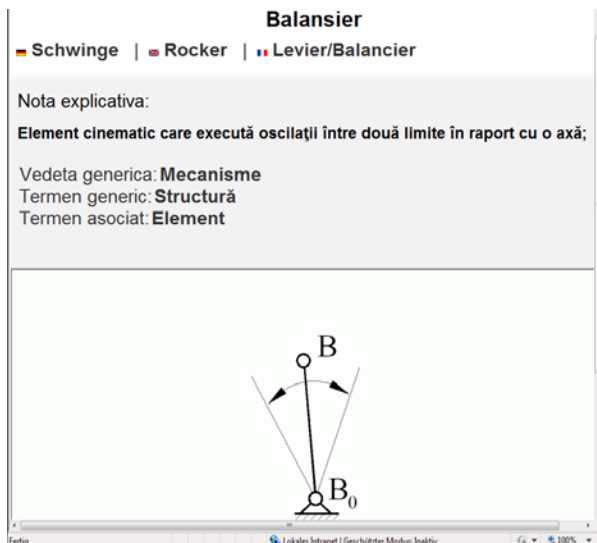

Fig. 8 Specific window of the descriptor "balansier" – "rocker"

From the specific windows of the descriptors it can be shown that they have the same associated term TA

"element" and also the generic term TG is "stuctură" – "structure". That is in concordance with the definition of the element in the field of mechanisms science – "each of the component parts constituting a mechanism" [8].

Example 3

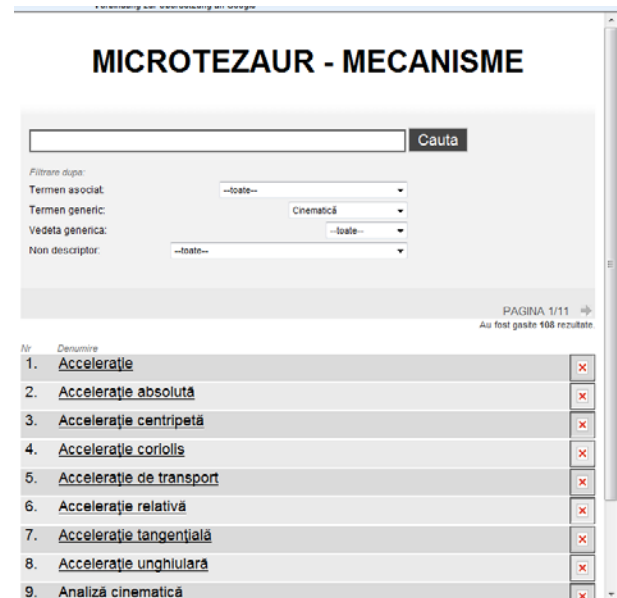If the advanced search is used for searching only in the field of the generic terms TG, for example on the

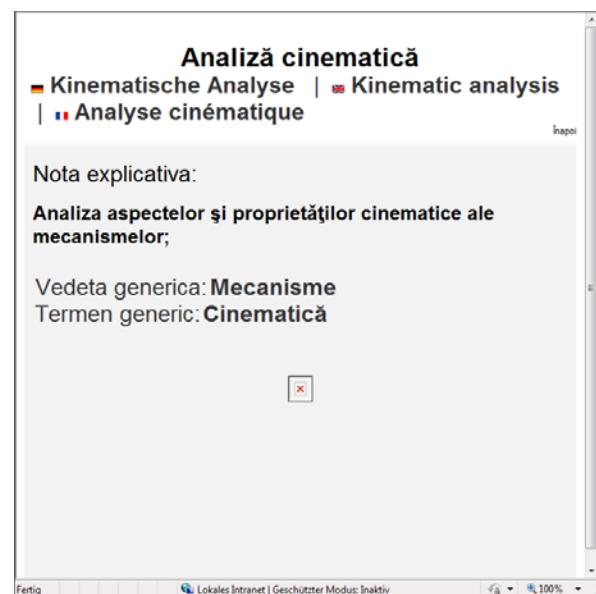Fig. 9 Interface of the μ-thesaurus database for the advanced search on the generic term "cinematic" – "kinematics"

Fig. 10 Specific window of the descriptor "analiză cinematică" – "kinematic analysis"

generic term "cinematică" = "kinematics" gives all the specific cinematic descriptors (108 indexing terms). These descriptors in the chosen example are: "acceleraţie" = "acceleration", "acceleraţie absolută" = "absolute acclelearation", "acceleraţie centripetă" = "centripetal acceleration", … , "analiză cinematică" = "kinematic analysis", etc.

These searching results are shown in Figure 9 in alphabetical order. It is to mention, that the advanced search for the generic terms TG is subordinated to the general headings VG and contains the specific chapters of mechanisms science. This observation can be shown in the Figure 10, where all the descriptors do not include in the specific windows the associated term TA.

Example 4

Some descriptors have synonyms, which are known as non-descriptors UP. To allow also the retrieval of these terms the non-descriptors are also included in the database, if these terms exist,. One example is the descriptor "batiu" = "frame", which has the synonym "bază" = "base". The specific window for the descriptor "frame" is shown in the Figure 11.
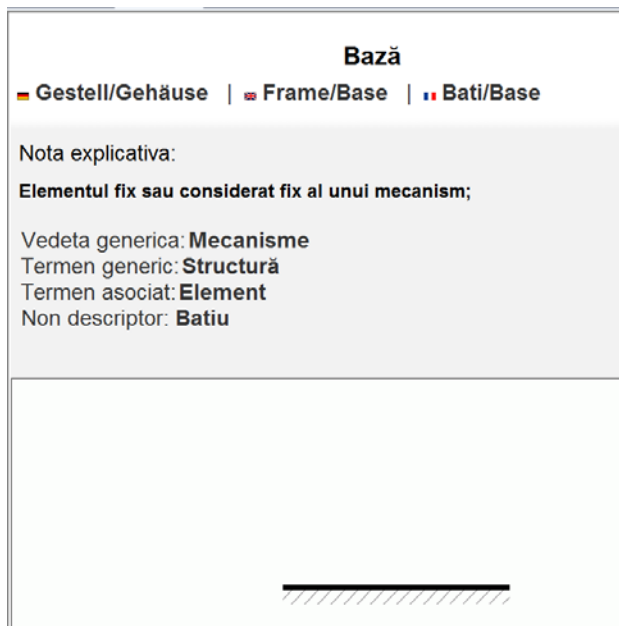

Fig. 11 Specific window of the descriptor "bază" – "frame"

VI. Conclusions

The multilingual illustrated μ-thesaurus in the science of mechanisms it is addressed to several types of users:
- librarian - indexer person specialized in indexing documents in the library, which use this tool for coordinated, consistent and unambiguous indexing;
- researcher as specialists in mechanisms science, to find some terms in his translation in different languages (English, German and French);
- students, which can access online that μ-thesaurus to be familiarized with the main concepts related to mechanisms science.

The use of the μ-thesaurus for the indexing of documents related to mechanism science allows a convergence between the indexing language and the user's search language, trying to retrieve sources of mechanism science.

The translation of the 385 descriptors of the μ-thesaurus database of mechanisms science indexing terms in English, German and French leads to overcoming the language barrier.

An important role has also the associated figure or picture with each descriptor of the database that facilitates rapid and intuitive information to the user.

In the case of coordinated indexing the user reaches faster the real information than in the case of free indexing.

This multilingual illustrated μ-thesaurus database will be used in framework of the European Project thinkMOTION [9] in order to support the retrieval in the Digital Mechanism and Gear Library [10].

References

[1] Library of Congress Subject Headings, 22nd edition, Washigton, D.C., Library of Congress, Cataloging Distribution Service, 1999.
[2] http://www.collectionscanada.gc.ca/csh/index-e.html
[3] Rameau: Répertoire d'Autorité-Matière Encyclopédique et Alphabé-tique Unifié, Montpellier, ABES, 1995.
[4] http://guiderameau.bnf.fr/html/rameau_0098.html
[5] http://www.nlm.nih.gov/mesh/
[6] STAS 12598-87. Metode de analiză a documentelor, de determinare a conţinutului lor şi de selectare a termenilor de indexare, Bucureşti, Institutul Român de Standardizare, 1987.
[7] http://130.15.85.212/terminology/TerminologyWeb/index.html.
[8] Terminology for the mechanism and machine science. Mechanism and machine theory, vol. 38 (2003), p. 597-605.
[9] http://www.thinkmotion.org
[10] http://www.dmg-lib.org