DESIGN OF AN OPEN FORMULAE FOR THE INTERDISCIPLINARY MULTILINGUAL TERMINOLOGICAL DICTIONARY FOR GEOINFORMATICS

J. J. Jachimski

University of Science and Technology in Krakow, Poland, Department of Photogrammetry and Remote Sensing Informatics
jjachimski@hotmail.com, jjachim@uci.agh.edu.pl

Commission VI

KEY WORDS: Photogrammetry, Remote Sensing, GIS, terminology, dictionary, Internet

ABSTRACT:

Fast development of computer science oriented on a capturing, processing and distribution of geoinformation (*geoinformatics*), which has applications in many technical and natural disciplines of science and technique, calls for urgent unification of professional terminology, for terminological dictionary.
The presented open formulae of creation of dictionary for geoinformatics, assumes the use of internet, to continuously present the content of the dictionary data base. That way the dictionary will be available for users, even though it will not yet be completed. The open formulae assumes also participation of volunteers in editorial work, under supervision of responsible editors, and interactive internet discussion and comments from the site of users.
The project organized by the Commission for Geoinformatics of Polish Academy of Science and Skills in Kraków, Poland, is in the initial stage. The experimental version of pertaining computer program is ready for use.

1. INTRODUCTION

The contemporary world develops under great influence of the computer science. Most of the every day information can be gained from the internet. Also special data, concerning geo-spatial information can be achieved that way.

Very many people use geoinformatics. The language used to describe the technique of collecting, processing and providing geo-spatial information becomes popular, not only among specialists of technical geo-spatial sciences, and specialists of natural geo-spatial sciences, but also among people who use geoinformatics to solve their every day problems.

Commission for Geoiformatics at the Polish Academy of Science and Skills (KG PAU) in Kraków, Poland, has been formed in 1998. That commission gathers scientists of many disciplines, such as geology, geophysics, environmental engineering and protection, mining, surveying and geodesy, geography, cartography, GIS, and, of course, photogrammetry and remote sensing, and computer sciences (*informatics*). All these specialists found out just at the beginning of their cooperation, that they do not speak the same technical and scientific language concerning geoinformatics. Finally it was decided to work out a dictionary, which could help people of geoinformatics from various disciplines to use identical vocabulary [1,3].

The work was split into thematic teams (fig.1), which working independently, but together with each other are to build final product: The Interdisciplinary Multilingual Terminological Dictionary for Geoinformatics, a very important project of KG PAU.
The problem of terminological dictionaries was in the ISPRS focus from years [4,5,6].
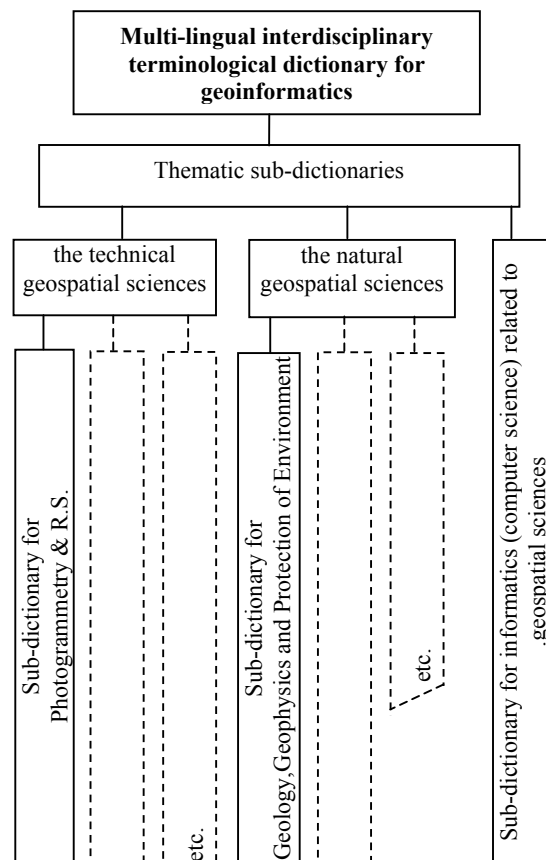

Fig. 1. This dictionary of geoinformatics is build of not limited number of thematic sub-dictionaries

Our discipline, Photogrammetry and Remote Sensing, is to prepare a sub-dictionary for itself. It is going to be something more, than only thematic sub-dictionary of the great dictionary for geoinformatics. Our dictionary, on the model of dictionaries for photogrammetry made in Poland before [7], will contain wide vocabulary set, the terms directly connected and only related to our thematic discipline. There are also another examples of polish thematical dictionaries covering photogrammetry and R.S. [1,8].

In opposite to our predecessors, the new dictionary will be build up in the open formulae, using internet. All the work stages will be public, available in internet. The dictionary users will have chance to comment the description of each term, and the whole work on the internet pages. To the editorial process will be admitted also volunteers, who, under care and in cooperation with the responsible editors will help to build the dictionary data base. Here mainly students preparing their master thesis, or PhD thesis, will try their hand at proper understanding even complicated technical and scientific terms belonging to photogrammetry, remote sensing and GIS. The *imaging geoiformatics* (the geoinformatics dealing with capturing, processing, and producing images), which covers, among others, the main part of photogrammetry, remote sensing and GIS, gathers many young enthusiasts around that discipline. They feel, that in future the Information Society will use imaging geoinformatics in a great extend, and will appreciate low cost of collecting data that way, and comfort of using the information in the easy to understand form of images.

2. THE OPEN FORMULAE OF THE EDITORIAL PROCESS

One of the main objectives of the project it is to develop a uniform terminology for such an interdisciplinary subject as geoinformatics. At the end of our project there should be presented a CD and printed book: the dictionary.

This aim is very clear, but it is not an easy goal, and not the only goal we want to achieve.

The geoinformatics develops fast, and simultaneously in many disciplines, sometimes quite far from each other. It is, therefore, rather impossible to stop independent development of several

thematic vocabularies for geoinformatics. The editorial activity, therefore, it is a permanent task. The only hope, the only way to progressively unify terminology, it is to provide a field of continuous linguistic consultation between the individuals of many fields of geoinformatics applications. Therefore it was decided, that the dictionary must be approachable in the internet, not only for passive, but also for the active use.

2.1. Need for Internet

The open editorial formulae, in our case, means that everybody at any time can consult the internet page with the actual version of dictionary on it. Everybody can use our dictionary, to support his professional editorial activity, but also can influence the proposals and interpretation of terms given in the dictionary. To achieve that very important goal, to each entry of the dictionary there is attached a computer window in which everybody can write his comments and proposals. The text of such a comment is stored in the database, and should help to accomplish work of dictionary editors. So, that commentary text is accessible always to everybody, as the all dictionary content. Such a situation can stimulate an internet discussion, and also can influence the linguistic decisions of people responsible for dictionary. It often happens that a term is used by practicing specialists with a new, additional meaning, which should be registered in dictionary. It also happens, though, that practicians invent a new term, which is rather horrible, and should be immediately replaced by another word, better fitting to the general rules of certain language. A prompt action on the side of editors, a proper comment to the comment of the dictionary user, can stop using the unwanted term. One should consider the well known rule, that also a professional language it is a living medium, and words, once spontaneously accepted – stay for long.

2.2. Editorial faces, tentative and final version

The open editorial formulae means also, that all the editorial activities are done in public. The term (an entry of dictionary), or a synonym, thematic description of meaning, and other associated texts, they are to be edited, and then verified and finally approved, as official dictionary version. All that editorial faces must be documented on the internet page of dictionary, and must be always accessible for the users.
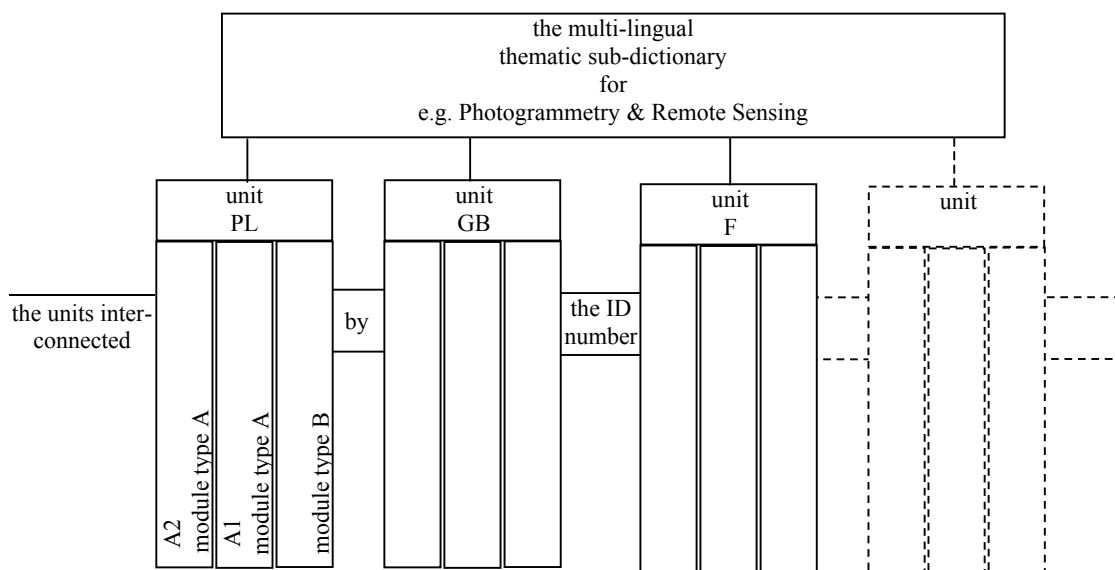

Fig.2. The multilingual units of any virtual thematic sub-dictionary. The number of languages is not limited by the software. A selected entry has identical ID in each of the languages. Each linguistic unit consists of two identical software modules type A (A1 for tentative and A2 for finally approved texts), and one module for links to attached articles pertaining to the entry (type B)..

In our virtual dictionary coexist two versions of each text: the tentative, and the final, official (fig.2). Together with them, under certain dictionary entry, there is also visible in internet all the pertaining discussion. Of course, in the final, printed version, in future, there will be published only final, official version of the texts. We assume, though, that the virtual, internet version of dictionary will never die, and all the documentation of any text creation process will be always accessible (not only for the historical studies of language development).

2.3. Editors

There are two types of dictionary editors: the responsible editors, and editors volunteers. The editor volunteer, rather amateur type in that activity, may work only in a team with the responsible editor. They create various texts for the dictionary, as a team. That team work product is introduced to dictionary as a tentative text, and it needs verification and approval by other authorized responsible editor. Also text prepared alone by a responsible editor is being introduced to dictionary as a tentative, and needs similar verification and approval by the other authorized responsible editor.

2.4. Students

The proposed organization opens the way for wide participation of amateurs in editorial activity. There is a great editorial potential in the university students, especially those preparing their master thesis, or PhD thesis. I think, that a text prepared for the dictionary together with the supervisor, the professor conferring the degree, and signed by both, could give satisfaction to authors. The editorial activity of wider group can greatly speed up the process of creation of dictionary. The participation of students in this process have a substantial influence not only on the dictionary content, but also in the first place, on the habit to use our dictionary in the everyday activity. It is not enough to create the dictionary. It will have a very little influence on the purity of the language, if it will not be in a popular use.

2.5. Inventory of thematic terms

The Multilingual Interdisciplinary Terminological Dictionary for Geoinformatics of PAU covers very wide thematic field. Specialists of many scientific and technical disciplines decided to work to create that dictionary. In the first step the actually used terms must be catalogued and described. It was decided to make that cataloguing in separate thematic sub-dictionaries, independently for each discipline, but not separately, in one editorial team split periodically to thematic or inter-thematic sub-teams. This thematic orientation may cause some deviation, in selection of dictionary entries, from terms purely related to geoinformatics, to the terms generally adequate to the particular thematic discipline (eg. such disciplines as: photogrammetry and remote sensing, geology, geography, GIS, cartography or other). Considering the virtual recording of texts – the memory space is practically not limited. Considering narrow specialization of thematic editorial teams – it would be wise to accept as the sub-directory entries also the terms that have no direct meaning for geoinformatics, but belong to the thematic discipline. That way in the sub-dictionaries can be registered also entries that would not pass the final selection, which will take place when selection and compilation of terms to the final version of the great geoinformatics dictionary will be executed. But the advantage for users of virtual version of the thematic sub-dictionary will be in more rich selection of terms for

themselves. This means, that the virtual thematic sub-dictionary is a very rich by-product of creation of the great geoinformatics dictionary.

2.6. Languages and linking the dictionaries

Each thematic dictionary, and after integration, also the final dictionary for geoinformatics, they can be edited in parallel in as many languages as required. The languages can be carried on in full extent of the dictionary texts, or only equivalents of terms in other languages may be introduced. The other texts in some languages can be added to dictionary at any chosen time. Also an another, new language can be introduced at any time to dictionary.

It would be easy to merge dictionaries edited separately in various languages (eg. in various countries). There must be, in both merged dictionaries, at least one common language (a matching language). Finding corresponding entries in this language in both dictionaries, gives a match to the entries, and other pertaining texts, in all languages present in dictionaries being linked.

3. THE LOGIC OF THE WORK OF SOFTWARE FOR THE VIRTUAL DICTIONARY

3.1. The data base records

Each entry of the Dictionary, and the associated texts are registered in the dictionary data base, in the data base modules.

Basic module type "A" of the dictionary software					
identification number of the main entry	main entry (the term)	entries for synonyms	short thematic description of the entry	examples of the proper and wrong use of the entries (the terms)	the links to the attached encyclopedia texts, describing more in detail the problems associated with the entry (the term), and links to attached copies of pertaining important publications

Fig.3. Basic module type "A" of software of the dictionary is shown above. The "Module type B" it is just one interactive window field designated for "remarks and discussion"

The basic module, called here the module type A, consists of the following records (fig.3):

- o the identification number (ID) of an entry
- o main entry (the term)
- o entries for the synonyms
- o short thematic description of the term
- o examples of the proper and wrong use of the term
- o the links to the associated texts, more in technical detail describing the term (encyclopedic report and pertaining articles)

The module A1 serves, in the first place, for registering the tentative version of the texts. After they are verified, corrected and approved, they are moved and registered in the module A2 (see fig 2).

The other module of the data base, which belongs to the term, it is module type B. That module it is just one record which serves for writing down the comments and a free discussion, concerning the entry.

This three modules form an unit, used for registering each entry (the term) in a one language. The corresponding entries in other languages are registered in the similar units of the data base. The corresponding entrees in all languages used in a certain dictionary (or sub-dictionary) are linked by a common identification number (ID).

3.2. The program usage

The program of the dictionary, just after it is opened, asks for identification of the leading language, and three other languages, which will be accessible during the session. The leading language it is the language of all the headings and texts (eg. the helps) which belong to the program rather then to the dictionary data base, and will be shown on the screen.

The English it is, by default, the language of the program when it opens.

All four languages must be selected, though. When the user wants to use only one or two languages, the other two positions shall repeat one of the already selected languages.

After the languages are decided, on the screen appear composition of windows shown on fig. 4. Here one can identify the language for entries, which will be written by him in future in window (1). To execute that one of the buttons (4-7) must be clicked, to get the word YES on it. In a result all the entries written in window (1) will be selected only from that data base. Corresponding terms and synonyms in this and other languages will automatically be shown in windows (8-15).

Now, using buttons (16-20), the other texts associated with the entry term, can be exposed in window (16). These texts, in a tentative or final version (20), are:

- o short thematic description of the term (button 17)
- o examples of the proper and wrong use of the term (button 18)
- o the links to the associated texts, more in technical detail describing the term (encyclopedic report and pertaining articles) (button 19)

All these texts will be taken from the data base of that language, which was chosen as a language for selecting the entrees.
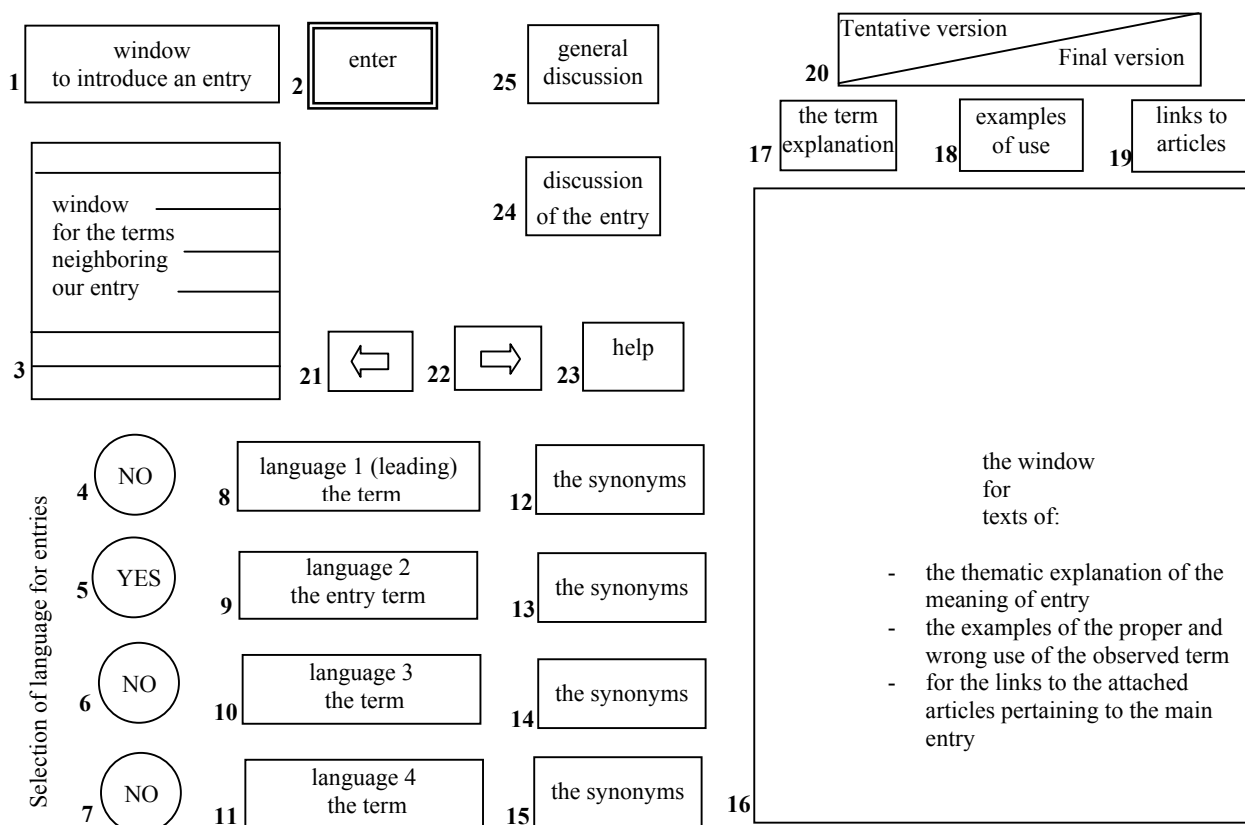

Fig.4. The main window of dictionary. One of the four preselected (4-7) languages can be chosen any moment to be the language for entries (1-2). The entry term and corresponding terms in other languages appear in windows 8-15. The explanation of meaning of main entry word, or examples of use, or the links to articles, are exposed in window 16 with help of buttons 17-20. Buttons 21 and 22 are for moving to previous or the next screen content.

Now, using buttons (16-20), the other texts associated with the entry term, can be exposed in window (16). These texts, in a tentative or final version (20), are:

- short thematic description of the term (button 17)
- examples of the proper and wrong use of the term (button 18)
- the links to the associated texts, more in technical detail describing the term (encyclopedic report and pertaining articles) (button 19)

All these texts will be taken from the data base of that language, which was chosen as a language for selecting the entrees.

That type of information may be present also in the data bases of the other three languages. To draw the user's attention to that, the buttons (4-7) change automatically the color from white to the bright green if in related data base are other texts. To approach these texts, that particular language must be selected as a language for entries. To do that one must select proper button (4-7), and to click it (the word YES will appear).

To use two or more typical settings of information composition on the screen, one can use buttons (21-22).

To read or to write some free comments concerning the entry, one can open the thematic discussion field by clicking the button (24). To read or write a free comment not to particular entry, but to the dictionary as a whole, the general discussion field can be opened by clicking button (25).

All the information written to the program data base are stored, as a documentation of stages of editorial work, or discussion and comments made by users. When it is too much information in the data base (specially that of the free discussion), it can be decided, which part of this information is approachable by a user in the first step, and which need to be specially called from the dictionary archives.

4. ORGANIZATION OF THE EDITORIAL WORK

The editors responsible for the dictionary have their editorial rights registered in the dictionary software. These rights mean that the responsible editor can introduce to the data base a new entry, and other texts associated with any term, but only as a tentative version. The rights are recognized by the program, when the editing person enters his password to the special window of the editing page on the screen.

As it was mentioned before, each responsible editor can work in a team. He can cooperate with any chosen person, a volunteer, but he takes responsibility for the texts prepared by the team members, under his supervision, and he gives final shape to the texts. Such texts are signed by the authors, but in the first place by the responsible editor.

The tentative version, after being approved, is moved to the module containing final versions of dictionary texts (fig.5). Only editors verifiers can approve, or change it, in consultation with the authors.

The editors and editors verifiers can edit the texts only in the languages for which they were authorized to do it for.

Each thematic sub-dictionary has a separate team of editors assigned to it.

There are portions of vocabulary which can appear in more than one thematic sub-dictionary. Those words can be edited separately for each sub-directory, or can be edited by a team created by the editors of both (or more) sub-dictionaries. We call such a team of editors – the inter-thematic team of the inter-thematic dictionary portion. It is advisable, that such a team

works in independent data base, which is on a certain stage compiled to adequate common data base.
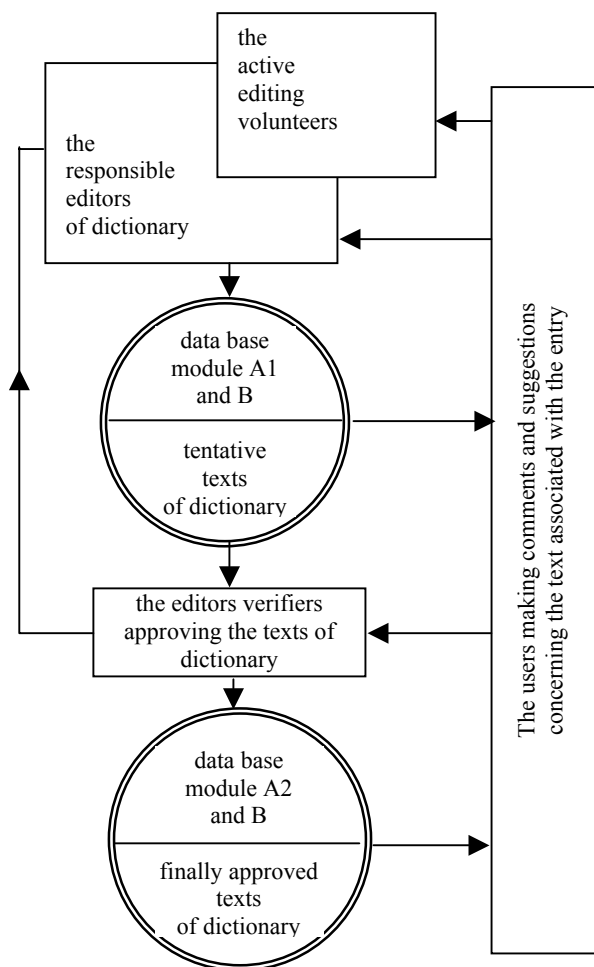

Fig. 5. The flowchart of the open formulae process of editing each separate entry of the dictionary via internet. The texts associated with each entry are prepared by editors, with a help of volunteers, and presented to public in internet as a tentative (A1). Only after verification by other editors they became the final dictionary texts (A2).

At the end of certain period of time, when the data bases are fairly complete, all the final texts are selected from the point of view if they belong thematically to the geoinformatics dictionary. That stage of editorial work is very important from the point of view of cooperation between various scientific technical and natural disciplines. After probably sharp terminological discussions (we have already quite an experience in it) finally the dictionary in a virtual version will be ready for the editing for printing. The advantage of the proposed open formulae of editing the dictionary is that at any stage of development the dictionary is available for users without limitation. The advantage from splitting the dictionary to several sub-thematic, and inter-thematic dictionaries, lays in easier organization of editorial work: the tempo of the work can be different in different groups, and nobody is dependent to anybody's work habits. It is easier to organize work in a small groups, and it is easier to create in the team members the sense of responsibility for the team work general results.

5. CONCLUSIONS

The open formulae of editorial work, the internet based formulae concerning the dictionary for geoinformatics, it is the way not only for faster collecting the information about the terminology, but also the way for mobilization people toward the precise and clear use of the technical language.

I express the hope, that the general rules of the open formulae, proposed here, will not be rejected, but in opposite, rather further developed by some other groups creating dictionaries for geoinformatics, or for photogrammetry, RS and GIS, in other countries. On the basis of international cooperation the members of ISPRS could be able to produce good dictionaries for many languages in a relatively short time. The cooperation within the EU could give an additional prompt to it.

For a good beginning I would be looking for at least bilateral cooperation.

6. BIBLIOGRAPHY

1. Dyskusja panelowa poświęcona zagadnieniom terminologicznym geoinformatyki (The panel discussion on geoinformatic terminology) edited by Janusz Kotlarczyk, Geoinformatica Polonica, 3:2001, PAU, Kraków

2. Gaździcki J.: Leksykon geomatyczny (The lexicon of geomatics), Polskie Towarzystwo Informacji Przestrzennej, Warszawa 2001

3. Jachimski J.: Wielojęzyczny Interdyscyplinarny Terminologiczny Słownik Geoinformatyczny Komisji Geoinformatyki PAU – zarys projektu pilotażowego (Multilingual interdisciplinary terminological dictionary for geoimformatics of The Geoinformatics Commission of Polish Academy of Science and Skills – an outline of the pilot project), Geoinformatica Polonica 2004, Kraków, in print

4. Lindig G., Bahr H.P., Sievers J.: ISPRS Multilingual Dictionary General Status and Progress of German Language Group, IAPRS, ISPRS, Washington 1992

5. Serge P. at al.: Constitution de la Partie Francaise du Dictionnaire Multilingue de la SIPT/ISPRS (Compilation of the French part of the Multilingual Dictionary from the ISPRS), IAPRS, ISPRS, Kyoto 1988, Vol.27, B8

6. Sitek Z.: Five-lingual Dictionary on Photogrammetry and Remote Sensing – Polish Edition, IAPRS, ISPRS, Washington 1992

7. Słownik pięciojęzyczny z zakresu fotogrametrii i teledetekcji z objaśnieniami w języku polskim" (The five-lingual Dictionary on Photogrammetry and Remote Sensing – with explanation in Polish), Zakład Fotogrametrii AGH, Kraków 1990. Edited by the team of authors under supervision of Zbigniew Sitek

8. Słownik podstawowych terminów używanych w teledetekcji (The dictionary of basic terms used in R.S.), PTG, Warszawa 1987, edited by the team of authors under editorial supervision of Jan Olędzki