DEKOMPOZICIJA BIBLIOGRAFSKIH REFERENCI U LOGIČKOM MODELU INFORMACIONOG SISTEMA U FIZICI ATOMSKIH SUDARA

DECOMPOSITION OF REFERENCE TEXTS IN LOGICAL MODEL OF INFORMATION SYSTEM IN ATOMIC COLLISION PHYSICS

V.Cvjetković, M.Bilbija, B.Marinković, * V. Bočvarski and V. Petrović

Faculty of Natural and Mathematical Sciences, Univ. of Kragujevac
****Institute of Physics, Belgrade, P.O.Box 68, 11080 Zemun***
E-mail: bratislav.marinkovic@phy.bg.ac.yu

Sadržaj *U ovom radu se razmatra problematika istraživanja i organizacije istraživanja sa tačke gledišta ekvivalentnog informacionog sistema koji obuhvata sve aktivnosti i podatke značajne za naučno istraživanje u oblasti fizike atomskih sudara. Jedna od glavnih teškoća je uvid i kako ostati u toku sa velikim i stalno rastućim brojem referenci koje su važne za naučno istraživanje. Predloženi informacioni sistem bi trebalo da služi kao alatka istraživaču, i da se koristi u svim fazama istraživanja. Jedna od glavnih karakteristika je dekompozicija članka u kategorije koje opisuju strukturu, osnovne karakteristike i sadržaj referenci, i omogućava pretragu po članovima kategorija. Razmatraju se modeli procesa i podataka predloženog informacionog sistema kao i predložena dekompozicija. Planira se buduća implementacija na WEB-u, što bi omogućilo široku dostupnost za sve zainteresovane iz date oblasti, kao i moguću saradnju na razvoju informacionog sistema.*

Abstract *The problem of research organizing and performance is discussed in this paper from the point of view of equivalent information system that comprise all the activities and data relevant for scientific research in the area of atomic physics collisions. One of the main obstacles is insight and being up to date with large and constantly increasing number of references that are relevant for scientific research. Proposed information system should serve as a tool for researcher, and would be used in all phases of the research. One of the main characteristics is article decomposition into categories that describe the structure, basic characteristics and contents of the reference, and enables the search by category terms. Process and data models of the proposed information system are discussed, as well as article decomposition. Future implementation on the WEB would enable wide availability for all interested in the field, and potential cooperation on information system development and distribution.*

1. INTRODUCTION

Organizing of any research implies insight into very broad and sometimes hardly available references that give the current achievements in the area of interest, and that can in some way initiate or inspire the future investigations. This problem is usually being solved so that every researcher keeps track of the researches in his scientific area of interest, which are published in different media. The aimed segment of the information gathered during the search is often of numerical type, and the information presented may be scattered through the text, or in the form of tables that are inserted in the text. Besides, no matter how good and systematic the work in following the events is, it is not rare case that in certain time period number of relevant information increase that much, that it is impossible to gain satisfactory insight if using traditional methods. Difficulties arise in disproportion when performing research using the method of synthesis of published results of other researches, because in that case data are spread over publications that differ not only in the kind of processes but in methodology also.

Up to now, much effort and work has been done in order to solve this problem. Search for data especially in the case of existing digital storage of bibliography units, is mainly performed by keywords so popular and extensively used on the Internet nowadays. Keyword search can be performed on full text, titles, abstracts, authors or some combination of mentioned. Search engines can utilize information about keywords and description stored in HTML meta tags. Keyword search can have many different varieties ranging from simple search to advanced search, possibility to combine few keywords with logical operators, to define exact phrase, use of the wildcards etc. Keyword search has very important disadvantage, and that is the presence of unwanted contents existing in found and ranked material – documents (out of the scope of interest) popularly called garbage. Reasons are that the human language generally speaking is a very rich one, and often-same word could be used (and is) with different meanings (homonyms), symbolic meanings, the presence of the word or even its frequency does not necessarily mean that the text is of any interest. As it is not yet possible for machine (search engine) to understand meaning of the text being searched, some heuristics have to be implemented. The first most important are the already mentioned position of the keyword in text, whether it is in the title, abstract, beginning, middle…The second is relative frequency comparing to other non key words. These heuristics are important when ranking the search

results. Better results (less garbage) could be obtained by additionally constraining the search to some area – geographic, scientific, etc, depending on what options are offered.

The method proposed in this contribution is a lot more general, and searching for data is just one of many aspects. It is based on information system that is intended to support research activity as a whole. Description of both models for processes and data of such information system is given here. As already mentioned, the search for scientifically important contents in wide literature is one of the basic activities in every research on one hand, and on the other hand it has to be significantly improved, so it is given very important role here. Offered solution for the search of scientifically important contents is complementary to already mentioned keyword search, so we do not try to improve the keyword search. Keyword search could be used here as a supplementary method. Search is based here on fragmentation of text into many important categories that characterize the field of investigation, in this case atomic collisions. Of course, principles are of universal kind, and are directly applicable to other areas. Main categories for text fragmentation of the contents of the article or book in this case, are in principle the same as independent entities that are given in the data model of the information system. Those entities are for instance author, publication, laboratory, projectile particle, target particle, process, method, interaction method, parameter, variable quantity, measured quantity, and others mentioned later in the text. Entity named publication contains data about texts and documents that were published and fragmented. These mentioned entities follow from the data flows in the process model.

For the time being, the necessary work on text fragmentation is human, more precisely the expert, as the document fragmentation must be done by people very well acquainted with the research area that the given document that is analyzed belongs to. It is also the human choice of the document that will be analyzed. Data resulting from the text fragmentation are stored in the tables of the database that ensures minimum redundancy of data and that the fragmentation of the given text has to be done only once, but that the results of fragmentation can be used unlimited number of times by various people. Fragmentation categories could be viewed as dimensions of text, and data for each category resulting from fragmentation as particular coordinate values, thus positioning the analyzed text into some kind of point in multidimensional space in which the document resides. With such concepts introduced, we could say that the search is to be performed by giving one or more "coordinate" values, and checking the set of "points" thus obtained. If all the coordinate values are supplied (except one that is being searched for, PUBLICATION for instance) then there is only one point, but in principle, there can be none, one or more documents corresponding to only one point. When speaking of coordinates of different kind (corresponding to different categories), one implicitly assumes that they are mutually connected with AND logical operator, although they could be connected using OR. In the case when more than one coordinate values of the same kind are given, they could be connected with OR logical operator, and in some cases they could be connected with AND, for instance there could be one or more authors of the article. In some other cases, different coordinates of the same kind could be mutually exclusive. Then quite various multidimensional areas or sets of points can be defined, some of them perhaps not containing and others containing documents.

As the entity corresponds to one dimension or coordinate, then entity instance corresponds or is what is called here coordinate value. Before main search, all entity instances that will be used must be specified. Specification of instances is made on the basis of known data for entity attributes, and could result in zero, one or more instances of the given entity. That way search for the documents would always be with existing instances that result from the done text fragmentation. One would then expect that every search with existing instances yield results, namely one or more documents. It may not always be the case, as the existing instances from various categories could be connected with logical operators into non-existing combination.

Second very important specificity of such approach is to realize that no text or document has to be in the underlying database, but just the data resulting from the text decomposition. Search would yield just the one or more particular articles or books that were already fragmented - analyzed. The useful information obtained that way is the reference to text, not the text itself. This information system is not to compete with the library, but is rather complementary one.

This approach requires considerable human work for document decomposition. To increase the effectiveness, many individuals could do decomposition of various documents in parallel – simultaneously, it could be made necessary requirement during their regular research work. In order for the results of their work to be mutually compatible, it is necessary to establish some standards. First of all, same things or concepts should be named the same, without using synonyms. That would require the starting list of category attributes and their values. Starting list would have of course the increasing tendency with time, as the new concepts are met or introduced in documents. Also, new categories or dimensions might be introduced, following the development in the research area. Adoption of new concepts should be made known to all the individuals doing the decomposition. In case of spreading the decomposition, for instance between institutions, then the choice of bibliography units for analysis should be made, in order to avoid potential duplication. Cooperation in the work of standardized text fragmentation, would enable direct exchange and use of the results obtained by others, thus increasing the effectiveness.

If this method is found useful, then the decomposition

categories and corresponding values for entity attributes in other areas of research could also be established, to help the research and researchers in other areas.

2. PROCESS MODEL

Logical model of every information system consists of process model and data model. Process model is that part of the overall informational model, in which the processes with their input and output flows are shown in the appropriate way. The role of these processes is, no matter on which decomposition level are presented, to transform the data received from their input data flows into data which are in their output data flows.

Results of designing the logical model of information system having the already mentioned capabilities in the introduction and applied in the physics of atomic collisions are presented earlier [1]. Method of Structural System Analysis [2] was used in that design. The aim was to perform the analysis of functions that happen in reality and that are to be computer supported, by proper methodological procedures. Process model is obtained from the analysis of the real events in scientific research. The so-called Data Flow Diagrams (DFD) of the various levels of decomposition of every process that is sensible to decompose is formed. On the top of decomposition pyramid is the most general DFD (so called context diagram) of the overall function that is subject of the computer support.

There is no direct procedure for obtaining the data model from the process model, and vice versa, but both follow from thorough analysis of the real events that are to be modeled for use in information system. Bellow it is presented in detail how it was achieved in the case of function "Research in the physics of atomic collisions (RPAC). Starting from context level, process model shows function RPAC as in figure 1.

In reality researchers perform the RPAC function, and publish the research results in some of the bibliographical forms. For obtaining necessary data about other researches in this area, researchers have the opportunity to search the existing database. Database also serves as storage for own research results, and research results of other researchers. As far as the data flows between RPAC function and bibliography as the element from it's surroundings are concerned, they are established in both directions, so that RPAC gives data about publishing, and also accepts published results from the earlier performed researches. In order for all of these to be possible, RPAC relays on database of research results for searching the already stored data, and also for storing the new data.

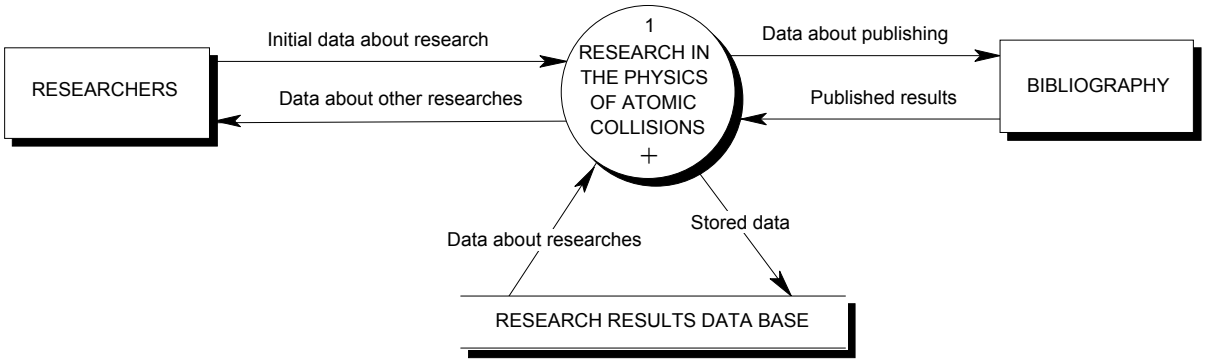

Figure 1. Context diagram of the function RPAC

By more detailed analysis it can be concluded that RPAC function can be derived according to following. Researchers, on the base of the set aim define criteria for searching and make insight into available data from the earlier performed researches. In this procedure, which can be iterative, initial conditions of the planned research are defined, i.e. preparation for computer-aid and for its organization is performed. After the organizing process is over, the next step is performing of the research in which the measurement of the observed phenomenon is made, i.e. some results are obtained. By the analysis of these results certain conclusions are drawn, that are the basis for their publishing. In the phase of publishing there are two activities going on. On one side relevant data and information are published in some of the bibliographical forms, and on the other preparation of data is made for their processing and storing in the information system database. Processing and storing in the database is also enabled for data that were published earlier. Accordingly, RPAC function can be decomposed on sub processes as shown in figure 2.

In order to get clearer picture of needs for data and information as well as for processes in which these are generated and used, each of the sub processes shown in figure 2 have to be further analyzed. (Full text with further decomposition on one or two additional levels, of the process model and descriptions of the sub diagrams, can be found on the following WEB address:

http://fizika.pmf.kragujevac.co.yu/prezentacija/istrazivanje/Ifas/Ifas.htm or http://147.91.204.212/prezentacija/istrazivanje/Ifas/Ifas.htm.

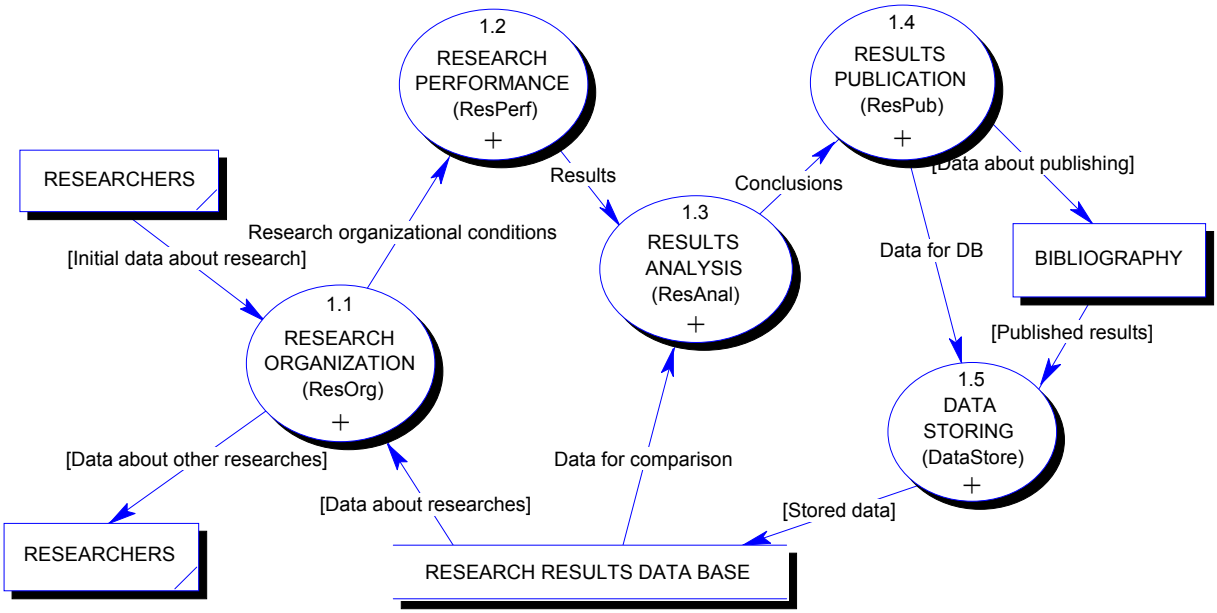

Figure 2. DFD of the decomposed RPAC function

3. DATA MODEL

Data model used here is relational and it is developed up to the level of recognition of all the entities, defining of identifiable and descriptive attributes of these entities, links between entities and properties of these links.

Namely, in heading 2 DFD are defined both on the context level and on the following two or three levels of decomposition. Those DFD represent processes, surroundings, data storages, and data flows between these elements. While the processes represent the basis for the design of the programs that implement data transformation from input to output data flows, the very data flows are the basis for defining of data model. Process model and data model, as already emphasized, make the two key components of every information system. In the following text on the basis of data flows analysis the look of the relational data model for support of the RPAC function is given. This is achieved by using the diagrams of the parts of the whole model with symbolic according to Integration Definition for Information Modeling (IDEF1X) [3] standards, and which relates to development of the data model. For that purpose method of logical design for data modeling and database design - "Entity-Relationship Modeling" ("ER" modeling) [4] was used. Procedure that leads to data model was divided into three phases. In the first phase independent (fundamental) entities are defined, in the second links between independent entities are analyzed and preliminary data model in which the links between entities besides others are allowed to be of the type many to many is made, and in the third dependent (associative) entities are defined with which all links of the type many to many are eliminated and that take in broader views on the overall database design of the RPAC function.

Contents of the data flows represent the data that are structured and organized as independent entities in the process model. Analysis of the data flows in DFD of the function RPAC data model, led to following independent entities and their characteristics: AUTHOR, LABORATORY, PUBLICATION, PARTICLE, PREPARATION, PROCESS, INTERACTION METHOD, METHOD, PARAMETERS, EXPERIMENT, VARIABLE QUANTITY, MEASURED QUANTITY, and RESULT.

EXPERIMENT is the organized activity through which the research in the physics of atomic collisions is performed. It is as the entity identified by the attribute "Experiment code" in the database and is characterized by identifiable attributes of the independent entities that relates to it as "parents" and other descriptive attributes. Identifiable attributes like this in the EXPERIMENT entity are qualified as outer keys. In that way EXPERIMENT entity as the central entity of the RPAC has not its own attributes, but is formed on the basis of "parent entities" attribute values.

EXPERIMENT

Experiment code
Projectile particle / Particle code
Target particle / Particle code
Projectile preparation / Preparation code
Target preparation / preparation code
Interaction method code
Laboratory code
Method code
Process code

Figure 3 Experiment entity

It is important to notice how are the entities PARTICLES and PREPARATION defined in the experiment in the sense of target and in the sense of projectile. It is achieved so that the keys of these entities "Particle code" and "Preparation code" appear twice, of course in different roles.

Entity experiment represents all kinds of research

activities, not only those using experimental equipment and measurement, so it is understood in the wider sense of meaning.

4. RELATIONAL DATA MODEL DIAGRAM

After defining of the entities and analysis of properties and characteristics of links between entities (verb like names, types, characteristics) diagram of relational data model is obtained as shown in figure 4. Links between entities of the type many to many are tolerated in the diagram. Links like that can be recognized in figure 4 as they have filled black circles on both ends. But, final relational data models mustn't have links like that, so they have to be converted into links of the type one to many. It is achieved by the method of defining the entity of associative type that is inserted in place of link of the type many to many.

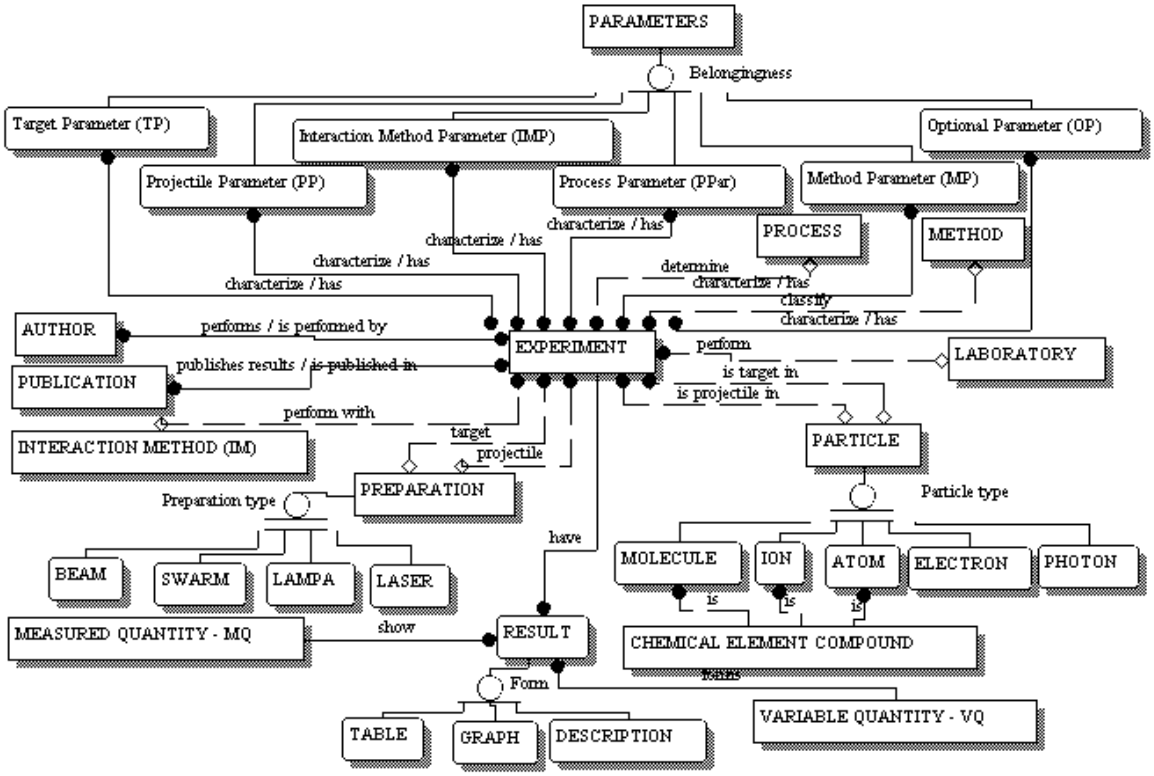

Figure 4 Relational data model diagram

5. CONCLUSION

Information system model in the physics of atomic collisions presented in this contribution was developed with the aim to enable fast and simple access to various data, which are necessary to every researcher, with the double role. On one side using this information system researcher can form own (local) information system in which by using the mentioned entities and their links he could update data of his own research (experiment) and thus having a tool for an easy and quick insight not just in the current state, but also in the previous events that where going on during the research. On the other side using the query for search that every researcher alone can define itself, enables selective access to information from the bibliography that is related for research (experiment) that is going on. As these both sides are just parts of the same information system, it is easy to set up connection between them.

Besides, as the decomposition of articles was performed on large number of attributes, the search is not performed on "free text", but on attributes, that enable much faster data selection process. It is important to emphasize that data systematization in RPAC enables its use in other areas of physics or chemistry.

Information system was built so that it follows the procedures and processes that the researcher goes through during consideration, organization and realization of his experiment. Process model presented in this paper is in fact the logical decomposition of the usual procedures of every researcher that is performed not just in physics of atomic collisions but also in majority of other areas in natural sciences. Even though here presented information system is connected by its attributes for physics of atomic collisions, it is obvious that its logical structure is easy adaptable to other scientific areas especially in natural sciences.

And also, data from information system like this can be made available for wider audience, using the WEB access, for search, or for new contributions.

REFERENCES

[1] M.Bilbija, B.Marinković, V.Cvjetković, V.Nikolić and V.Bočvarski, "*Logical model of information system*

in atomic collision physics" Proc. 19th SPIG, Zlatibor 1998, Contributed Papers, Eds. N.Konjević, M.Ć'uk and I.R.Videnović, pp.143-6

[2]. De Marco "*Structured Analysis and System Specification"*, Yourdon Press, 1978.

[3]. Standard "*Integration Definition for Information Modeling (IDEF1X)"*, Federal Information Processing Standards Publications (FIPS PUBS) are issued by the National Institute of Standards and Technology after approval by the Secretary of Commerce pursuant to Section 111(d) of the Federal Property and Administrative Services Act of 1949 as amended by the Computer Security Act of 1987, Public Law 100-235. **1993 December 21**

[4] Chen P.P "*The Entity-Relationship Model - Towards a Unified View of Data"* ACM TODS, Vol. 1 No. 1, Mar. 1976.

YU INFO 2002

Simpozijum o računarskim naukama
i informacionim tehnologijama

ZBORNIK APSTRAKATA

Kopaonik, 11-15.03.2002. god.

YU iNFO 2001

Simpozijum o računarskim naukama
i informacionim tehnologijama

ZBORNIK APSTRAKATA

PROGRAMSKE OBLASTI SIMPOZIJUMA

- **internet**
- **multimedija**
- **informacioni sistemi**
- **razvoj softvera i alati**
- **računarska simulacija**
- **računarske mreže i telekomunikacije**
- **računarski hardveri**
- **sistemi za podršku odlučivanja**
- **veštačka inteligencija i ekspertni sistemi**
- **e-vlada i e-poslovanje**
- **primenjena informatika**

Kopaonik, 11-15.03.2002. godine

Izdavač:
YU INFO asocijacija, 11000 Beograd, Paunova 9

Za izdavača:
prof. dr Miodrag Ivković

Odgovorni urednik:
mr Dušan Korunović

Redakcija:
Programski odbor

Tehnička obrada:
Mile Ž. Ranđelović, dipl. ing.

Štampa: "Sven", Niš
Tiraž: 400 primeraka

YU INFO 2002

Organizatori :

YU INFO asocijacija u saradnji sa Univerzitetom u Nišu

Programski odbor:

Prof. dr Borivoje Lazić, ETF Beograd
Prof. dr Božidar Radenković, FON Beograd
Prof. dr Borko Furht. Florida Atlantic University
Prof. dr Branimir Djordjević, ELF Niš
Prof. dr Branislav Lazarević, FON Beograd
Prof. dr Bratislav Milovanović, ELF Niš
Prof. dr Dragan Perović, ETF Podgorica
Prof. dr Dragan Domazet, MNTR Beograd
Prof. dr Dragić Banković, PMF Kragujevac
Mr Dušan Korunović, Univerzitet u Nišu
Prof. dr Dušan Surla, PMF Novi Sad
Prof. dr Djordje Paunović, ETF Beograd
Prof. dr Miroslav Trajanović, MF Niš
Mr Mihajlo Djorović, Telekom Srbije
Prof. dr Milorad Stanojević, SF Beograd
Prof. dr Milutin Čupić, FON Beograd
Prof. dr Miodrag Ivković, MNTR Srbije
Prof. dr Pavle Mogin, FTN Novi Sad
Prof. dr Slobodanka Djordjević-Kajan, ELF Niš
Prof. dr Siniša Borović, VJ
Prof. dr Veljko Milutinović, ETF Beograd
Prof. dr Vladan Deverdžić, FON Beograd
Prof. dr Zoran Jovanović, ETF Beograd
Prof. dr Zora Konjović, FTN Novi Sad
Prof. dr Živko Tošić, ELF Niš
Dr Predrag Petrović, naučni savetnik, IRITEL
Prof. dr Zoran Bojković, SF Beograd

Organizacioni odbor:

Prof. dr Miodrag Ivković, predsednik
mr Dušan Korunović, sekretar
Prof. dr Božidar Radenković
Prof. dr Bratislav Milovanović
Prof. dr Veljko Milutinović

Pokrovitelj:

Ministarstvo za nauku, tehnologije i razvoj Republike Srbije

Generalni sponzor: CISCO SYSTEMS

Sponzor CD-ROM-a: SAGA

Sponzori:

- CISCO SYSTEMS
- DELL
- DATALAB
- MDS
- Siemens
- Ei INFORMATIKA Niš
- Privredna komora Niš
- SAGA
- ERICSSON
- INFORMATIKA Beograd
- EUNET Yugoslavia
- POGLED Niš
- SBS
- PAKOM Nis
- MediaNis
- Niška banka
- MICROSOFT
- INFOTEHNA
- TELEFONIJA
- TEHNICOM
- JISA Info
- CORES
- IRVAS Niš
- Nacionalne kuće "SURI"
- EKSTREM Niš

Sadržaj

INTERNET

MULTIMEDIJA

INFORMACIONI SISTEMI

RAZVOJ SOFTVERA I ALATI

RAČUNARSKA SIMULACIJA

RAČUNARSKE MREŽE I TELEKOMUNIKACIJE

RAČUNARSKI HARDVERI

SISTEMI ZA PODRŠKU ODLUČIVANJA

VEŠTAČKA INTELIGENCIJA I EKSPERTNI SISTEMI

E-VLADA I E-POSLOVANJE

PRIMENJENA INFORMATIKA

Informacioni sistemi

DEKOMPOZICIJA BIBLIOGRAFSKIH REFERENCI U LOGIČKOM MODELU INFORMACIONOG SISTEMA U FIZICI ATOMSKIH SUDARA

DECOMPOSITION OF REFERENCE TEXTS IN LOGICAL MODEL OF INFORMATION SYSTEM IN ATOMIC COLLISION PHYSICS

V. Cvjetković, M. Bilbija, B. Marinković, * V. Bočvarski, V. Petrović
Faculty of Natural and Mathematical Sciences, Univ. of Kragujevac
**Institute of Physics, Belgrade, P.O.Box 68, 11080 Zemun*
E-mail: bratislav.marinkovic@phy.bg.ac.yu

U ovom radu se razmatra problematika istraživanja i organizacije istraživanja sa tačke gledišta ekvivalentnog informacionog sistema koji obuhvata sve aktivnosti i podatke značajne za naučno istraživanje u oblasti fizike atomskih sudara. Jedna od glavnih teškoća je uvid i kako ostati u toku sa velikim i stalno rastućim brojem referenci koje su važne za naučno istraživanje. Predloženi informacioni sistem bi trebalo da služi kao alatka istraživaču, i da se koristi u svim fazama istraživanja. Jedna od glavnih karakteristika je dekompozicija članka u kategorije koje opisuju strukturu, osnovne karakteristike i sadržaj referenci, i omogućava pretragu po članovima kategorija. Razmatraju se modeli procesa i podataka predloženog informacionog sistema kao i predložena dekompozicija. Planira se buduća implementacija na WEB-u, što bi omogućilo široku dostupnost za sve zainteresovane iz date oblasti, kao i moguću saradnju na razvoju informacionog sistema.

The problem of research organizing and performance is discussed in this paper from the point of view of equivalent information system that comprise all the activities and data relevant for scientific research in the area of atomic physics collisions. One of the main obstacles is insight and being up to date with large and constantly increasing number of references that are relevant for scientific research. Proposed information system should serve as a tool for researcher, and would be used in all phases of the research. One of the main characteristics is article decomposition into categories that describe the structure, basic characteristics and contents of the reference, and enables the search by category terms. Process and data models of the proposed information system are discussed, as well as article decomposition. Future implementation on the WEB would enable wide availability for all interested in the field, and potential cooperation on information system development and distribution.

DEFINING UML MODEL TEMPLATE FOR WEB APPLICATION DEVELOPMENT

Dušan Veljković [1], Svetlana Cvetanović [2], Marko Veljković [1]
[1]Markom Systems Int'l - Toronto, [2]Univerzity of Niš

This document presents rationale for defining UML Model template that can be implemented for the developing Web applications. First part emphasizes issues of using raw UML modeling, and specifics of the Web application architecture. Using "Portfolio Tracker Project" as an example, the document gives a walk-through of the customization of the standard RUP modeling methodology and shows diagrams and structures as modeling results received in the different project stages. Also listed are the Web application projects for the different market segments, to which the Project team has applied this UML Template.

INDEKS AUTORA I KOAUTORA