



JOINT INSTITUTE FOR NUCLEAR RESEARCH
Veksler and Baldin laboratory of High Energy Physics

FINAL REPORT ON THE SUMMER STUDENT PROGRAM

*Development of a software package for
organizational activities of the Joint Institute
for Nuclear Research*

Supervisor:

Igor Stanislavovich
Pelevanyuk

Student:

Anna Ilina, Dubna State
University

Participation period:

July 01 – August 25

Dubna, 2019

Contents

Introduction.....	3
IndicoAbstract.....	3
JINRex.....	6
Methods	8
IndicoAbstract.....	8
JINRex.....	9
Results	11
IndicoAbstract.....	11
JINRex.....	13
Summary	18
Acknowledgments	19
References.....	20

Introduction

The administrative and organizational activities of JINR include a lot of various tasks, among which there is the creation of a book of abstracts for conferences which are constantly organized at the Institute as well as the organization and accounting of excursions for different types of auditory. During SSP I developed a software package to automate these tasks. It includes two products: IndicoAbstract: generator of Indico's Books of Abstracts and JINRex: JINR System of organization and accounting of the excursions. Let us dwell on each of the products in more detail.

IndicoAbstract

Book of abstracts contains a large number of pages stylized in a special way and containing information about the conference, speakers and their abstracts (fig. 1, fig. 2). 6-7 people manually wrote the book, continuous work took about 20 man-hours.

Creating such a book has a number of nuances:

1. The input file is generated by the Indico web service in .xml format (fig. 3).
2. All authors write Affiliations in their own way, which must be brought to one standard (fig. 4).
3. If the two authors have the same affiliation, it must be indicated once in the document.
4. Each author must have an affiliation index (fig. 5).
5. Each Speaker must have an email index (fig. 5).
6. Abstracts must be divided into sections.
7. In each section, Abstracts must be sorted by the names of the authors.
8. Each article does not allow the use of two or more languages. All paragraphs should be written in one language (fig. 3).
9. The style and formatting of the document should be identical to the document written by hand.
10. The generated document must have an editable format.

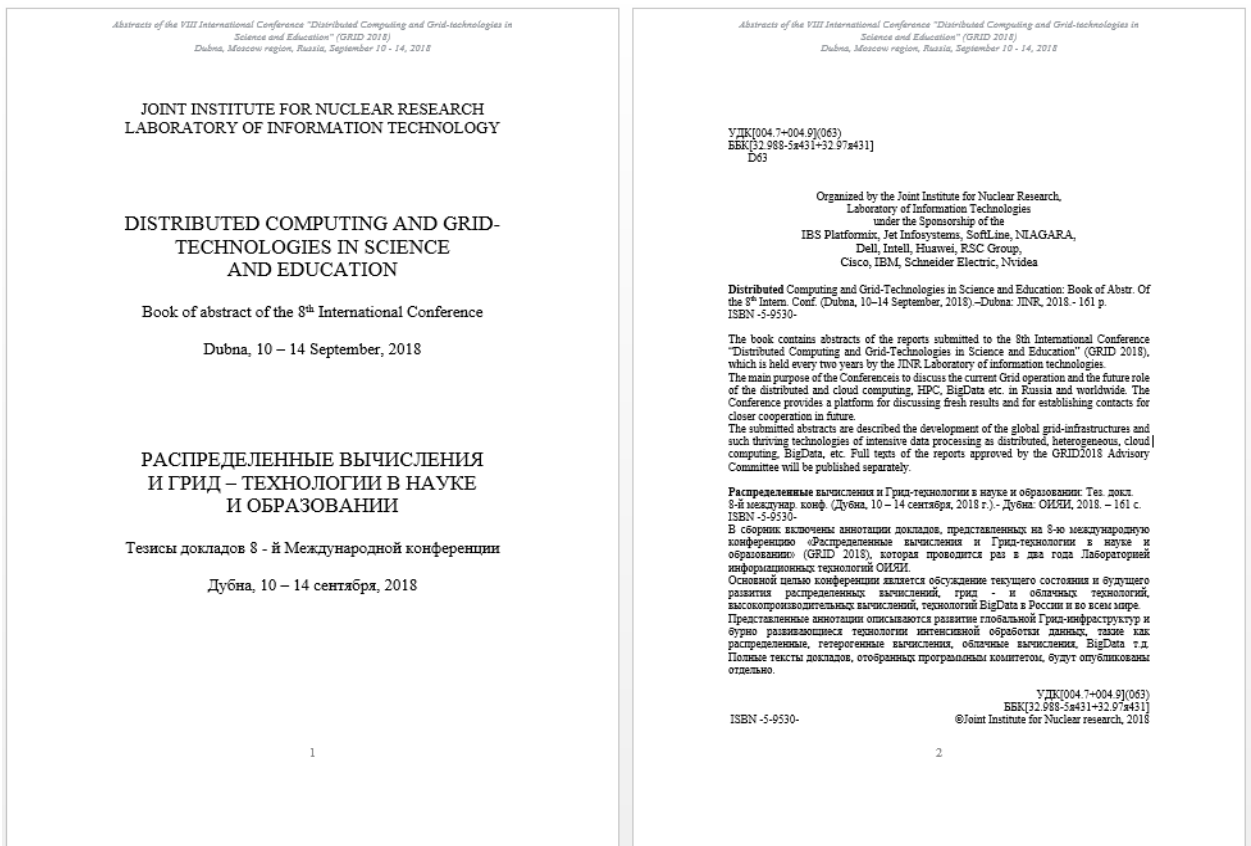


Figure 1. Book of abstracts frontpages

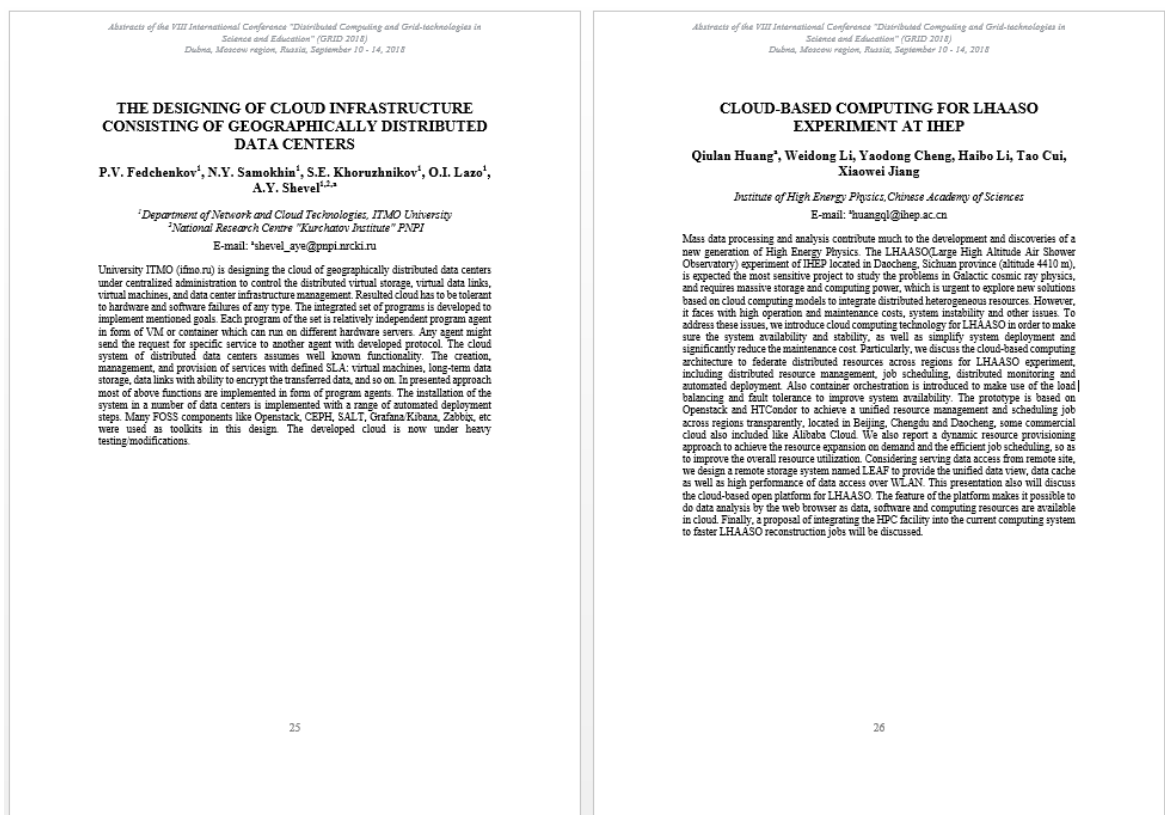


Figure 2. Book of abstracts content



```

▼<AbstractBook>
  ▼<Conference>
    The 8th International Conference "Distributed Computing and Grid-technologies in Science and Education" (GRID 2018)
  </Conference>
  ▼<abstract>
    <Id>180</Id>
    ▼<Title>
      Грид и облачная инфраструктура дата-центра Института Физики НАН Азербайджана
    </Title>
    ▼<Content>
      Основными направлениями развития дата-центра Института физики НАН Азербайджана являются грид и облачные технологии. Пользуясь тем что грид-сегмент дата-центра интегрирован в инфраструктуру EGI/WLCG в качестве грид-сайта уровня Tier3, пользователи дата-центра получают возможность участвовать в таких международных проектах как ATLAS (CERN). Сотрудничество с международными научными центрами ОИЯИ, ЦЕРН в области информационных технологий способствует эффективному развитию дата центра, а также помогает решать задачи пользователей в таких научных областях, как физика высоких энергий, физика твердого тела и т. д.
    </Content>
    ▼<field id="content">
      Основными направлениями развития дата-центра Института физики НАН Азербайджана являются грид и облачные технологии. Пользуясь тем что грид-сегмент дата-центра интегрирован в инфраструктуру EGI/WLCG в качестве грид-сайта уровня Tier3, пользователи дата-центра получают возможность участвовать в таких международных проектах как ATLAS (CERN). Сотрудничество с международными научными центрами ОИЯИ, ЦЕРН в области информационных технологий способствует эффективному развитию дата центра, а также помогает решать задачи пользователей в таких научных областях, как физика высоких энергий, физика твердого тела и т. д.
    </field>
    <field id="summary"/>
    ▼<PrimaryAuthor>
      <FirstName>Aleksey</FirstName>
      <FamilyName>Bondyakov</FamilyName>
      <Email>aleksey@jinr.ru</Email>
      <Affiliation>JINR (Joint Institute For Nuclear Research)</Affiliation>
    </PrimaryAuthor>
    ▼<Speaker>
      <FirstName>Aleksey</FirstName>
      <FamilyName>Bondyakov</FamilyName>
      <Email>aleksey@jinr.ru</Email>
      <Affiliation>JINR (Joint Institute For Nuclear Research)</Affiliation>
    </Speaker>
    <ContributionType>Sectional reports</ContributionType>
    <Track/>
  </abstract>
▼<abstract> ...
  
```

Figure 3. Example of the input file generated by Indico web service in .xml format

```

Joint Institute for Nuclear Research:Joint Institute for Nuclear Research
JINR:Joint Institute for Nuclear Research
LIT JINR:Joint Institute for Nuclear Research
LIT, JINR:Joint Institute for Nuclear Research
JINR LIT:Joint Institute for Nuclear Research
JINR, LIT:Joint Institute for Nuclear Research
FLNR JINR:Joint Institute for Nuclear Research
JINR (Joint Institute For Nuclear Research):Joint Institute for Nuclear R
Jinr:Joint Institute for Nuclear Research
jinr:Joint Institute for Nuclear Research
Joint Institute for Nuclear Research, Flerov Laboratory of Nuclear Reacti
ЛИТ ОИЯИ:Объединенный Институт Ядерных Исследований
  
```

Figure 4. Variation in writing Joint Institute of Nuclear Research and bringing it to one standard

THE ATLAS PRODUCTION SYSTEM PREDICTIVE ANALYTICS SERVICE: AN APPROACH FOR INTELLIGENT TASK ANALYSIS

Mikhail Borodin¹, Dmitry Golubkov², Alexei Klimentov³, Mikhail Titov^{4,a}

¹The University of Iowa (US)

²Institute for High Energy Physics

³Brookhaven National Lab

⁴National Research Centre «Kurchatov Institute»

E-mail: ^amikhail.titov@cern.ch

The second generation of the Production System (ProdSys2) of the ATLAS experiment (LHC, CERN), in conjunction with the workload management system - PanDA (Production and Distributed Analysis), represents a complex set of computing components that are responsible for defining, organizing, scheduling, starting and executing payloads in a distributed computing infrastructure. ProdSys2/PanDA are responsible for all stages of (re)processing, analysis and modeling of raw and derived data, as well as simulation of physical processes and functioning of the detector using Monte Carlo methods. The prototype of the ProdSys2 Predictive Analytics (P2PA) is an essential part of the growing

Figure 5. Affiliation and email indexes

JINRex

Conducting excursions is a very important task for JINR, as it helps people better understand the significance of the Institute's activities and provide a wider understanding of the processes that occur when solving a specific problem facing it. Automation of organizational activities for excursions will greatly facilitate the work of people in this area.

System of organization and accounting of the excursions should be a web service that satisfies the following conditions:

1. The system should have a human-friendly interface.
2. It should have a user authorization and authentication system.
3. Each authenticated user should be assigned to one of the roles:
 - a. Organizer – the person can create an application for an excursion.
 - b. Incharge – the person who is responsible for conducting the excursion at the specific facility.
 - c. Guide – the person who conducts excursions.
4. Each facility should have its own set of areas where excursions can be held.
5. The form for sending an application for an excursion have to include the following fields:
 - a. Desired facility.
 - b. Desired set of areas from those that are attached to the facility.
 - c. Full user's name from the list of Organizers.
 - d. Full user's name from the list of Guides.
 - e. Full user's name from the list of Incharges.
 - f. Occasion excursion.

- g. Date excursion.
 - h. Time period excursion.
 - i. Language excursion.
 - j. Type of auditory excursion.
 - k. Number of participants.
 - l. Age of the participants.
6. The system should have a mechanism for viewing all requests for excursions.
 7. The system should have a mechanism for changing data in the application for the excursion.
 8. Each of excursions should have Status field which can take one of two values: Confirmed and Unconfirmed. When creating or changing an application for an excursion the status should change to Unconfirmed until all participating persons (Organizator, Guide and Incharge) confirm these changes. Each person must receive a notification. After confirmation, the status of the application for the excursion should change to Confirmed.

Methods

IndicoAbstract

The method of generating documents based on templates in the .docx format. It is based on the fast process of creating a text template with variables that will be automatically replaced with the necessary data using special libraries (fig. 6).

To implement document generation using this method, it is necessary to create a software product with the appropriate library. Such a product should be able to read information from the input file and use the library to replace it with the corresponding variables. The disadvantages include the fact that if the corresponding library does not work correctly, the final document will also be incorrectly generated (there is a dependence on third-party software developers). [1] This approach is most practical for the generation of documents with large volumes of information, which must be issued in a special way, since:

1. Allows you to achieve your goal.
2. Supports the WYSIWYG property (What you see is what you get). [2]
3. Does not require writing .tex- or .html-views, unlike the LaTeX system. [3]
4. The final document has the same set of styles for formatting the text that was used in the template, which guarantees full compliance of the generated document with the standard.
5. The final document saves all notes made in the template (fig. 6).
6. The final document has an editable .docx format, which guarantees the possibility of editing information in case of its change.
7. A .docx format document can be easily exported to a .pdf document and immediately sent to print.

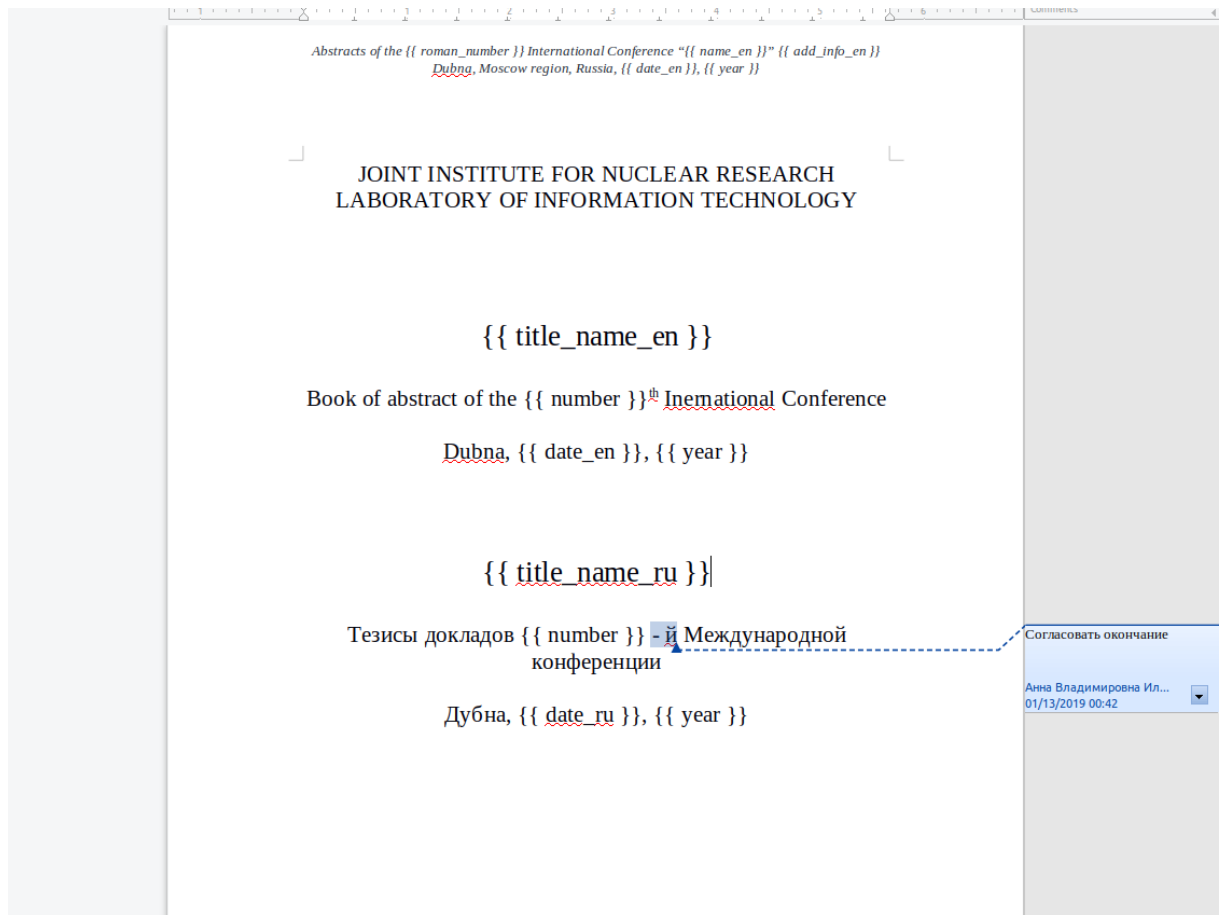


Figure 6. Example template with variables for generating book of abstracts

The implementation includes the following components:

1. Programming language Python v3.6. [4]
It works well with strings and has a ton of third-party libraries, as well as rich documentation.
2. Library xml.etree.cElementTree for working with XML files in pure Python which is used when processing large files. [5]
3. Library python-docx for creating Microsoft Word (.docx) files. [6]
4. Library python-docx-template (docxtpl) v0.6.3 for modifying Microsoft Word (.docx) files. [7]

The file "matches.csv" is used to bring the same but differently written affiliations to one standard. It contains information of the following content (fig. 4):

<wrong affiliation>:<right affiliation>

JINRex

The implementation includes the following components:

1. Programming language Python v3.6. [4]

Python is an easy to learn, powerful programming language. Python's elegant syntax and dynamic typing, together with its interpreted nature, make it an ideal language for scripting and rapid application development in many areas on most platforms. [8]

2. Django.

It is a high-level Python Web framework that encourages rapid development and clean, pragmatic design. Built by experienced developers, it takes care of much of the hassle of Web development, so you can focus on writing your app without needing to reinvent the wheel. Moreover, it's free and open source. [9]

3. Bootstrap v3.1.1.

Bootstrap is the most popular HTML, CSS, and JS framework for developing responsive, mobile first projects on the web. [10]

4. jQuery v1.11.0.

jQuery is a fast, small, and feature-rich JavaScript library. It makes things like HTML document traversal and manipulation, event handling, animation, and Ajax much simpler with an easy-to-use API that works across a multitude of browsers. [11]

5. Crispy Forms v1.0.0.

Django-crispy-forms is a Django application that lets you easily build, customize and reuse forms using your favorite CSS framework, without writing template code and without having to take care of annoying details. [12]

6. DataTables v1.10.18.

DataTables is a plug-in for the jQuery Javascript library. It is a highly flexible tool, built upon the foundations of progressive enhancement, that adds all of these advanced features to any HTML table. [13]

Results

IndicoAbstract

The generator is a console software product corresponding to all a priori representations.

The style and formatting of the generated document are fully identical to the document written by hand.

The source code and result document are available at: <https://github.com/trnkv/IndicoAbstract> (GNU General Public License v3.0).

The pictures below show the result of the program in Linux (fig. 7 – 9).

```
anna@anna-X751LD:~/Документы/git/IndicoAbstract$ python3 main.py
1. Parsing all abstracts from XML
2. The following affiliations are unknown. Please add them to CSV file with standards.
Viktorovich
Leading researcher
University of Sofia, Bulgaria
ITEP - NICKI
Research Assistant, Plekhanov Russian University of Economics, Moscow, Russia
Scientific Center in Chernogolovka, National Research University Higher School of Economics
Professor, Higher School of Economics
Alexandrovna
ISDCT SB RAS
Institute for System Dynamics and Control Theory of Siberian Branch of Russian Academy of Sciences
МГУ им М.В. Ломоносова
Ph.D. student
Dmitrievich
NRNU "MEPhI"
Brookhaven National Laboratory
Lomonosov Moscow State University
Германович
V. I. Il'ichev Pacific Oceanological Institute FEB RAS
Science Center in Chernogolovka
Applied Physics Institute Irkutsk State University
RCF "Computer Science and Control" RAS
Dubna State University
JOINT INSTITUTE FOR NUCLEAR RESEARCH
National Research Nuclear University MEPhI (NRNU MEPhI)
Ph.D.
St. Petersburg state Electrotechnical University
CPPM-IN2P3-CNRS
Saint Petersburg Electrotechnical University
Project Center ITER
INP BSU
Stankin
PNPI
Alekseevich
GÉANT/CESNET
IMPB RAS - the Branch of KIAM RAS
University of Sofia / JINR
student
Karlsruhe Institute of Technology (KIT)
IMPB RAS - the Branch of KIAM RAS
Institute for Information Transmission Problems of the Russian Academy of Sciences
Institute for Information Transmission Problems RAS
ТНФН
```

Figure 7. Warning information about affiliations which are not contained in the matches.csv file

```

=====
More than one language is used in abstract with Id: 180
Email to contact Speaker: ['aleksey@jinr.ru']
Speaker's name: ['Aleksey Bondyakov']
{'Author0_Affiliation': 'LATIN',
 'Author0_Name': 'LATIN',
 'Content': 'CYRILLIC',
 'Title': 'CYRILLIC'}

-----
More than one language is used in abstract with Id: 188
Email to contact Speaker: ['riviera-molto@mail.ru', 'evgenia.popova1397@gmail.com']
Speaker's name: ['eugene shchetinin', 'Evgenia Popova']
{'Author0_Affiliation': 'LATIN',
 'Author0_Name': 'LATIN',
 'Author1_Affiliation': 'LATIN',
 'Author1_Name': 'LATIN',
 'Content': 'CYRILLIC',
 'Title': 'CYRILLIC'}

-----
More than one language is used in abstract with Id: 193
Email to contact Speaker: ['svpoluyan@gmail.com']
Speaker's name: ['Sergey Poluyan']
{'Author0_Affiliation': 'LATIN',
 'Author0_Name': 'LATIN',
 'Content': 'CYRILLIC',
 'Title': 'CYRILLIC'}

-----
More than one language is used in abstract with Id: 197
Email to contact Speaker: ['a.iluhin@nordavind.ru']
Speaker's name: ['Андрей Илюхин']
{'Author0_Affiliation': 'LATIN',
 'Author0_Name': 'CYRILLIC',
 'Content': 'CYRILLIC',
 'Title': 'CYRILLIC'}

-----
More than one language is used in abstract with Id: 207
Email to contact Speaker: ['far334_oxid@mail.ru']
Speaker's name: ['Maxim Bystrov']
{'Author0_Affiliation': 'LATIN',
 'Author0_Name': 'LATIN',
 'Content': 'CYRILLIC',
 'Title': 'CYRILLIC'}

-----
More than one language is used in abstract with Id: 218
Email to contact Speaker: ['tsvetkov.vp@tversu.ru']
Speaker's name: ['Виктор Цветков']

```

Figure 8. Warning information about abstracts which have more than one language

```

More than one language is used in abstract with Id: 331
Email to contact Speaker: ['ivan-tatarinov@yandex.ru']
Speaker's name: ['Ivan Tatarinov']
{'Author0_Affiliation': 'LATIN',
'Author0_Name': 'LATIN',
'Author1_Affiliation': 'LATIN',
'Author1_Name': 'LATIN',
'Author2_Affiliation': 'LATIN',
'Author2_Name': 'LATIN',
'Author3_Affiliation': 'LATIN',
'Author3_Name': 'LATIN',
'Author4_Affiliation': 'LATIN',
'Author4_Name': 'LATIN',
'Content': 'CYRILLIC',
'Title': 'CYRILLIC'}
=====
WARNING: too few words in content of abstract with Id= 180 : 65 words
WARNING: too few words in content of abstract with Id= 183 : 49 words
WARNING: too few words in content of abstract with Id= 184 : 47 words
WARNING: too few words in content of abstract with Id= 189 : 91 words
WARNING: too few words in content of abstract with Id= 190 : 91 words
WARNING: too few words in content of abstract with Id= 191 : 76 words
WARNING: too few words in content of abstract with Id= 193 : 63 words
WARNING: too few words in content of abstract with Id= 194 : 75 words
WARNING: too few words in content of abstract with Id= 200 : 94 words
WARNING: too few words in content of abstract with Id= 201 : 61 words
WARNING: too few words in content of abstract with Id= 204 : 26 words
WARNING: too few words in content of abstract with Id= 209 : 49 words
WARNING: too few words in content of abstract with Id= 210 : 24 words
WARNING: too few words in content of abstract with Id= 213 : 72 words
WARNING: too few words in content of abstract with Id= 219 : 96 words
WARNING: too few words in content of abstract with Id= 227 : 82 words
WARNING: too few words in content of abstract with Id= 229 : 79 words
WARNING: too few words in content of abstract with Id= 232 : 61 words
WARNING: too few words in content of abstract with Id= 234 : 44 words
WARNING: too few words in content of abstract with Id= 236 : 87 words
WARNING: too few words in content of abstract with Id= 237 : 94 words
WARNING: too few words in content of abstract with Id= 238 : 68 words
WARNING: too few words in content of abstract with Id= 239 : 72 words
WARNING: too few words in content of abstract with Id= 244 : 82 words
WARNING: too few words in content of abstract with Id= 247 : 63 words
WARNING: too few words in content of abstract with Id= 252 : 73 words
WARNING: too few words in content of abstract with Id= 254 : 67 words
WARNING: too few words in content of abstract with Id= 255 : 54 words

```

Figure 9. Warning information about abstracts in the content of which the number of words goes beyond the established boundaries (less than 100 and more than 360)

JINRex

At the moment, the system is a nearly complete (not implemented user notification function) web service that satisfies a priori representations and has a small data set in a fully functioning database.

The source code is available at: <https://github.com/trnkvj/jinrex> (GNU General Public License v3.0).

Below are screenshots showing the results of the web service (fig. 10 – 17).

System of organization and accounting of excursions in JINR

Please login to see this page.

Username*

tmkv

Password*

.....

Login

[Lost password?](#)

Figure 10. The login page of JINRex

System of organization and accounting of excursions in JINR

Hello, Anna Ilina!

Logout

+ Fill out an application for an excursion

View all excursions

Figure 11. The main page of JINRex

System of organization and accounting of excursions in JINR

Hello, Anna Ilina!

Logout

Send a new request for the organization of the excursion

Select facility*

.....

Select desired facility

Select areas*

- Photos
- Museum
- MICC
- MICC CR
- Lectures

Organizer*

test testov

Enter the name of the excursion organizer

Select desired guide*

Ivan Ivanov

Select desired guide for this excursion

Incharge*

Petr Petrov

Responsible for the excursion

Occasion excursion*

Enter occasion of excursion (e.g. JEMS 12 etc.)

Date excursion*

Time period excursion*

Enter time period of excursion

Language excursion*

Enter language of excursion

Auditory excursion*

Enter auditory of excursion

Enter the number of participants*

Enter count of excursion participants

Enter the age of the participants*

Enter age of excursion participants

SEND

Figure 12. Form for sending an application for an excursion

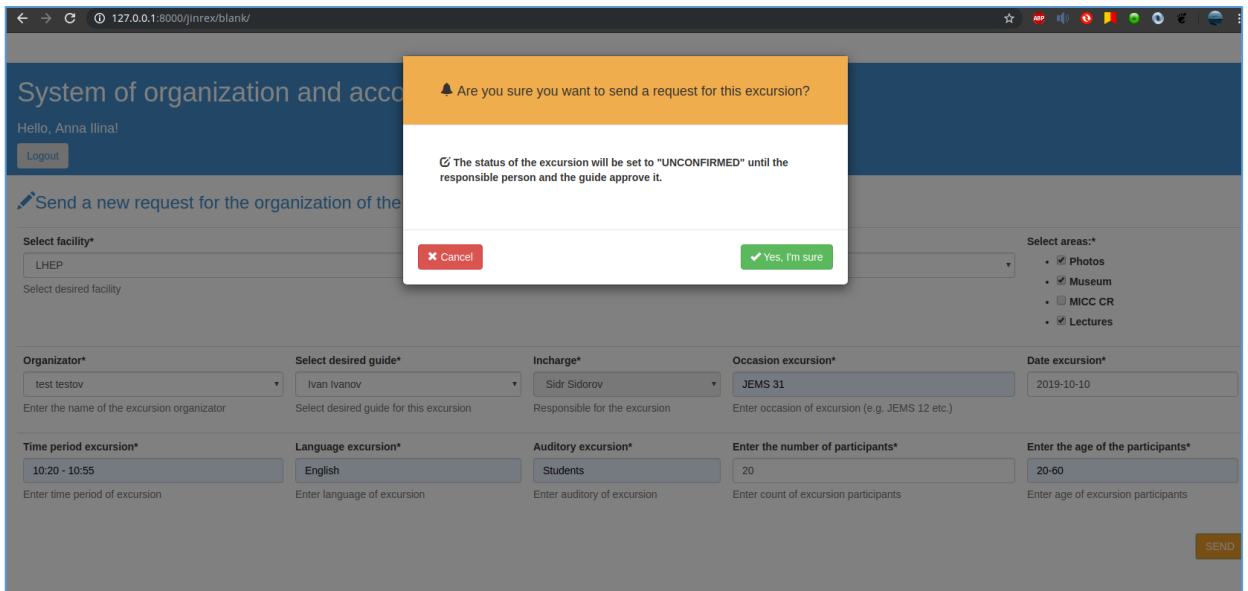


Figure 13. Dialog window to confirm the sending of the application

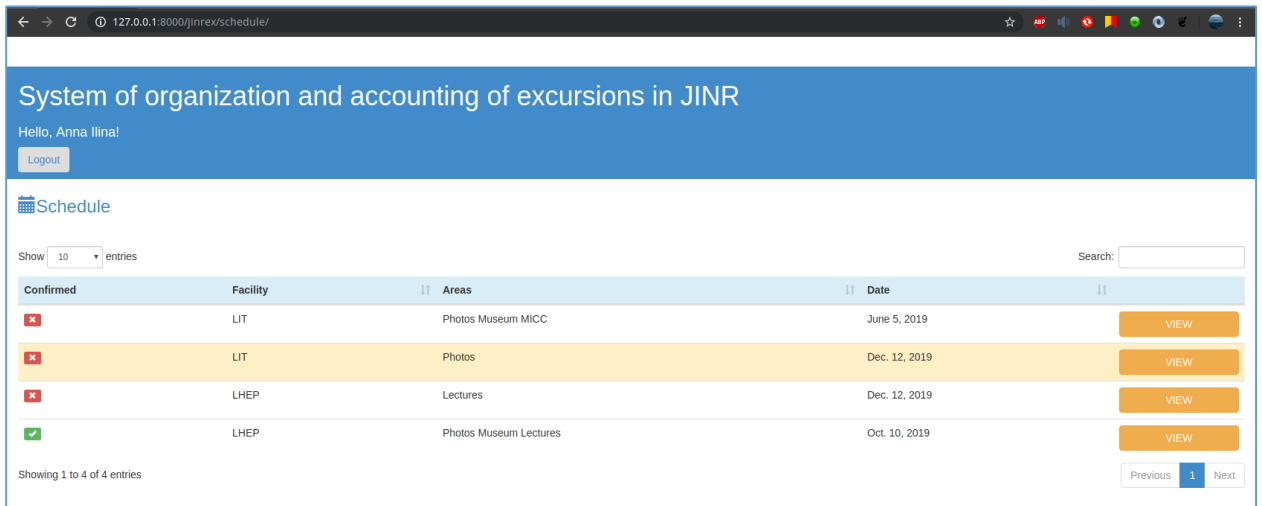


Figure 14. The page for displaying all excursion applications available in the database

System of organization and accounting of excursions in JINR

Hello, Anna Ilina!

Logout

To change a field just enter a new value and click CHANGE EXCURSION

Unconfirmed ✕

Select facility*
LIT
Select desired facility

Select areas:*
 Photos
 Museum
 MICC

Organizator*
test testov
Enter the name of the excursion organizer

Select desired guide*
Ivan Ivanov
Select desired guide for this excursion

Incharge*
Petr Petrov
Responsible for the excursion

Occasion excursion*
1111111111
Enter occasion of excursion (e.g. JEMS 12 etc.)

Date excursion*
2019-12-12

Time period excursion*
10:00-12:00
Enter time period of excursion

Language excursion*
English
Enter language of excursion

Auditory excursion*
Delegates
Enter auditory of excursion

Enter the number of participants*
10
Enter count of excursion participants

Enter the age of the participants*
20-60
Enter age of excursion participants

GO BACK CHANGE EXCURSION

Figure 15. Form to view and change data of unconfirmed excursion

System of organization and accounting of excursions in JINR

Hello, Anna Ilina!

Logout

To change a field just enter a new value and click CHANGE EXCURSION

Confirmed ✓

Select facility*
LHEP
Select desired facility

Select areas:*
 Photos
 Museum
 MICC CR
 Lectures

Organizator*
test testov
Enter the name of the excursion organizer

Select desired guide*
Ivan Ivanov
Select desired guide for this excursion

Incharge*
Sidr Sidorov
Responsible for the excursion

Occasion excursion*
JEMS 31
Enter occasion of excursion (e.g. JEMS 12 etc.)

Date excursion*
2019-10-10

Time period excursion*
10:00-12:00
Enter time period of excursion

Language excursion*
English
Enter language of excursion

Auditory excursion*
Students
Enter auditory of excursion

Enter the number of participants*
20
Enter count of excursion participants

Enter the age of the participants*
20-60
Enter age of excursion participants

GO BACK CHANGE EXCURSION

Figure 16. Form to view and change data of confirmed excursion

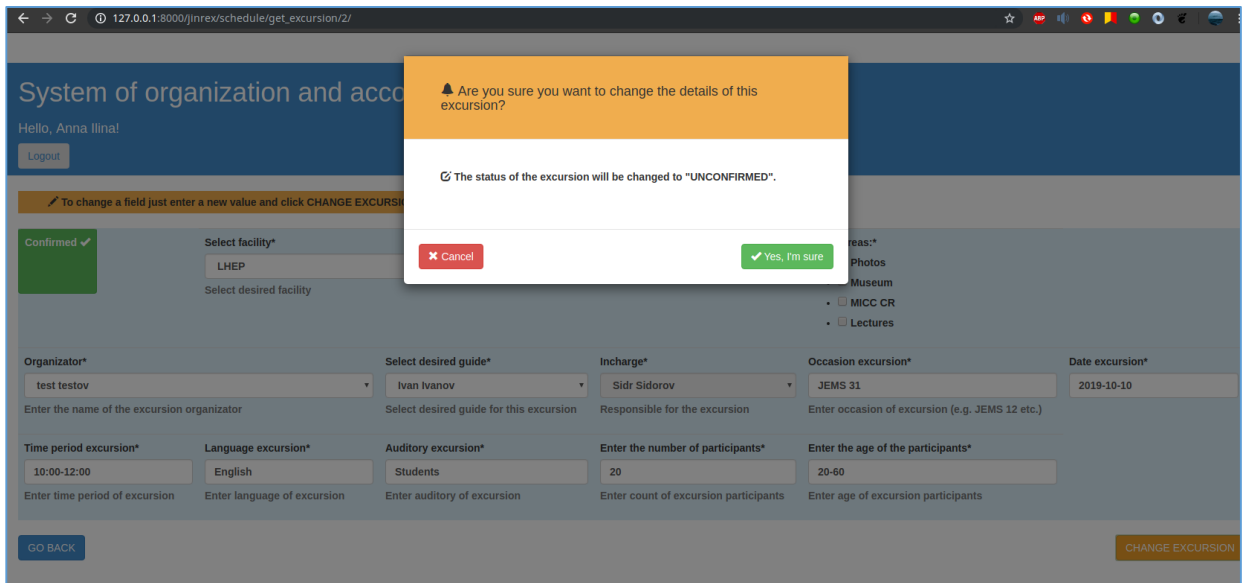


Figure 17. Dialog window to confirm the changing of the application

Summary

During SSP I gained a lot of knowledge and skills in working with a whole stack of various information technologies. In the process of developing software products, I improved my skills in analyzing systems, and also saw the software development process as a whole.

In the course of the study, it was found that the analogues of IndicoAbstract create books of abstracts only in LaTeX format, which indicates the uniqueness of created generator.

In the future, it is planned to add a graphical interface and functions for generating other documents (for example, badges for speakers) and to implement user notification function in JINRex system and фдџц install the system on the server.

Acknowledgments

I am very grateful to my supervisor Igor S. Pelevanyuk for interesting tasks, invitation and given opportunity to take part in this especially interesting and valuable student program. Thanks for assistance provided in some organization moments, for his support, patience, kindness, considerateness for me and high professionalism in many areas of information technologies.

Special thanks to Summer Student Program organizers for providing the opportunity to being a part of such prestigious scientific institution as JINR. I am especially grateful to JINR and SSP organizers in particular for the financial support and for assistance provided in solving all the issues.

Also I would like to thank my university teaching staff. Thanks to professor, doctor of technical sciences, director of the Laboratory of Information Technologies Vladimir V. Korenkov for given recommendation and for suggested idea to participate in SSP.

References

1. Automate the generation of docx documents by template using python and docxtpl - IT Blog. URL: <http://tut-admin.ru/2017/05/31/автоматизируем-генерацию-docx-документов/>.
2. WYSIWYG - Wikipedia. – URL: <https://ru.wikipedia.org/wiki/WYSIWYG>.
3. TeX - a system for typesetting texts with formulas. – URL: http://www.wikiznanie.ru/ru-wz/index.php/TeX_-_система_для_вёрстки_текстов_с_формулами.
4. Welcome to Python.org. – URL: <https://www.python.org/>.
5. Python: working with XML files and the xml.etree.ElementTree module. – URL: <https://rtfm.co.ua/python-rabota-s-xml-fajlami-i-modul-xml-etree-elementtree/>.
6. python-docx — python-docx 0.8.10 documentation. – URL: <https://python-docx.readthedocs.io/en/latest/>.
7. docxtpl · PyPI. – URL: <https://pypi.org/project/docxtpl/>.
8. The Python Tutorial — Python 3.7.4 documentation. – URL: <https://docs.python.org/3/tutorial/index.html>.
9. The Web framework for perfectionists with deadlines | Django. – URL: <https://www.djangoproject.com/>.
10. Bootstrap · The world's most popular mobile-first and responsive front-end framework. – URL: <https://getbootstrap.com/docs/3.3/>.
11. jQuery. – URL: <https://jquery.com/>.
12. Installation — django-crispy-forms 1.0.0 documentation. – URL: <https://django-crispy-forms.readthedocs.io/en/latest/install.html>.
13. DataTables | Table plug-in for jQuery. – URL: <https://datatables.net/>.