




ITMO UNIVERSITY

Linking E-Learning Ontology Concepts with NLP Algorithms

Dmitry Mouromtsev, Fedor Kozlov, Liubov Kovriguina and Olga Parkhimovich
Laboratory ISST @ ITMO University, St. Petersburg, Russia

16th FRUCT Conference, 27-31 October 2014, Oulu, Finland

Introduction

- E-learning ontology describes relations between educational resources (course, module, lecture, task, term),
- Use semantics to make education materials reusable and flexible,
- We need to provide tools for tutors to improve their courses  have to extract and interlink relevant educational materials (e.g., associate terminology in lectures and tests) and gather analytics

Some details...

1. The developed E-Learning ontology allows to create interdisciplinary relations between courses \Rightarrow a system may advice to repeat terms not only from the current course, but from the previous courses also (e.g., term “Vector” has links to courses “Vector algebra” and “Physics”);
2. When terms are linked to tasks (via “hasTerm” property), it is possible to get the data like
 - which terms present only in lectures, but not in tasks (\Rightarrow teachers add new tasks),
 - which terms present only in tasks, but not in lectures (\Rightarrow teachers add new explanatory content),
 - statistics;
3. A way to get balanced educational content.

Goal

- To develop the test ontology,
- To convert tasks from XML to RDF,
- To extract terms from tasks,
- To map tasks with ontology domain terms via extracted terms.

ECOLE: Front-end

- Ontology-based e-learning system
- User friendly interface
- Based on Django framework

ECOLE Fedor Kozlov Sign out

My courses +Add Profile

The theory of graphs
NRU ITMO
Modules: 8, Video lectures: 52, Tests: 1

No image
Fedor Kozlov
NRU ITMO

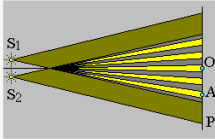
ECOLE Fedor Kozlov Sign out

Physics
NRU ITMO
Modules: 2, Video lectures: 7, Tests: 1

Information
Lectures
Tests
Practices
Literature
Terms
Ontology
Feedback

Тесты / Test Of Interference And Diffraction Frenel

Task 3. Finish



The figure shows plane P with light travelling from sources S1 and S2. The fringe width is the spacing between

- No right answer.
- neighboring high and low intensity peaks.
- neighboring high intensity peaks or low intensity peaks.
- zero-order maximum and the first intensity minimum.
- zero-order maximum and the point of observation.

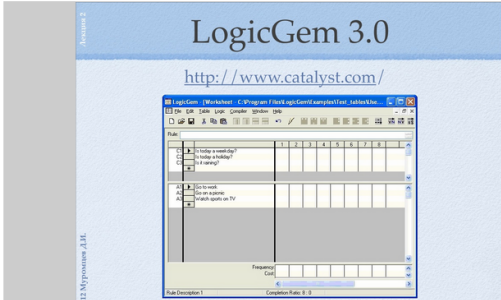
Поиск...

ECOLE Fedor Kozlov Sign out

Intelligent Systems
NRU ITMO
Modules: 1, Video lectures: 5, Tests: 1

Information
Lectures
Tests
Practices
Literature
Terms
Ontology
Feedback

Lectures / Overview of models and tools for knowledge representation



LogicGem 3.0
<http://www.catalyst.com/>

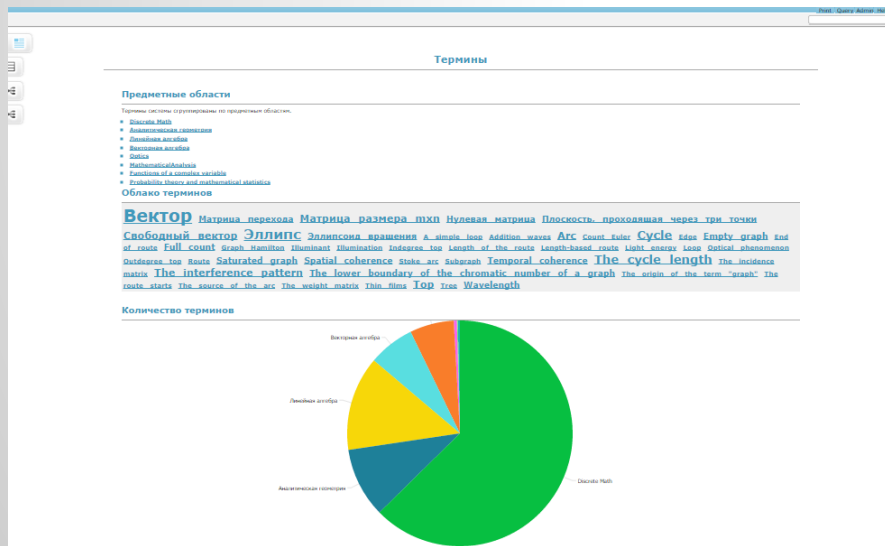
© 2010 MyIntelligence J.S.L.

15

sdulfmo.ru eShare 15 / 23

ECOLE: Back-end

- Collection of educational materials from different open resources (DBpedia BNB),
- Analytics tools,
- Based on the Information Workbench platform.



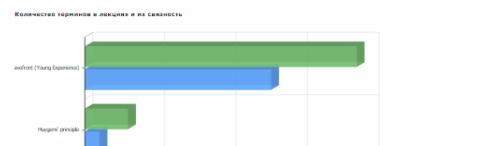
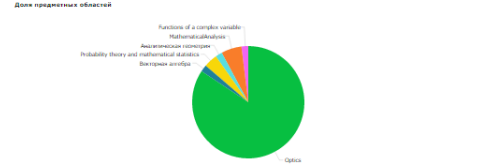
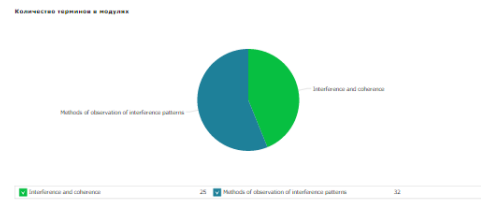
The screenshot shows the ECOLE front-end interface. It features a search bar at the top and a navigation sidebar. The main content area displays a search result for 'Linear algebra'. A 'Contents' box lists links for 'Linear algebra', 'Vector spaces', 'Examples', and 'Sources'. Below this, there is a section titled 'Количество терминов в модуле' (Number of terms in the module) with a pie chart showing the distribution of terms across different modules. The largest module is 'Discrete Math'.

In [mathematics](#), a **null vector** is an element of a [vector space](#) that in some appropriate sense has zero magnitude. In a [vector space](#) with a [bilinear form](#), a vector that is self-[orthogonal](#) (i.e. on which the bilinear form is zero) is referred to as a null vector. In a [seminormed vector space](#), it refers to a vector with zero seminorm. In contrast, the term **zero vector** refers to the unique [additive identity](#) of the vector space. In contexts in which the only null vector is the zero vector (such as [Euclidean vector space](#)) or where there is no defined concept of magnitude, *null vector* may be used as a synonym for *zero vector*.

- Contents
- [Linear algebra](#)
- [Vector spaces](#)
- [Examples](#)
- [Sources](#)

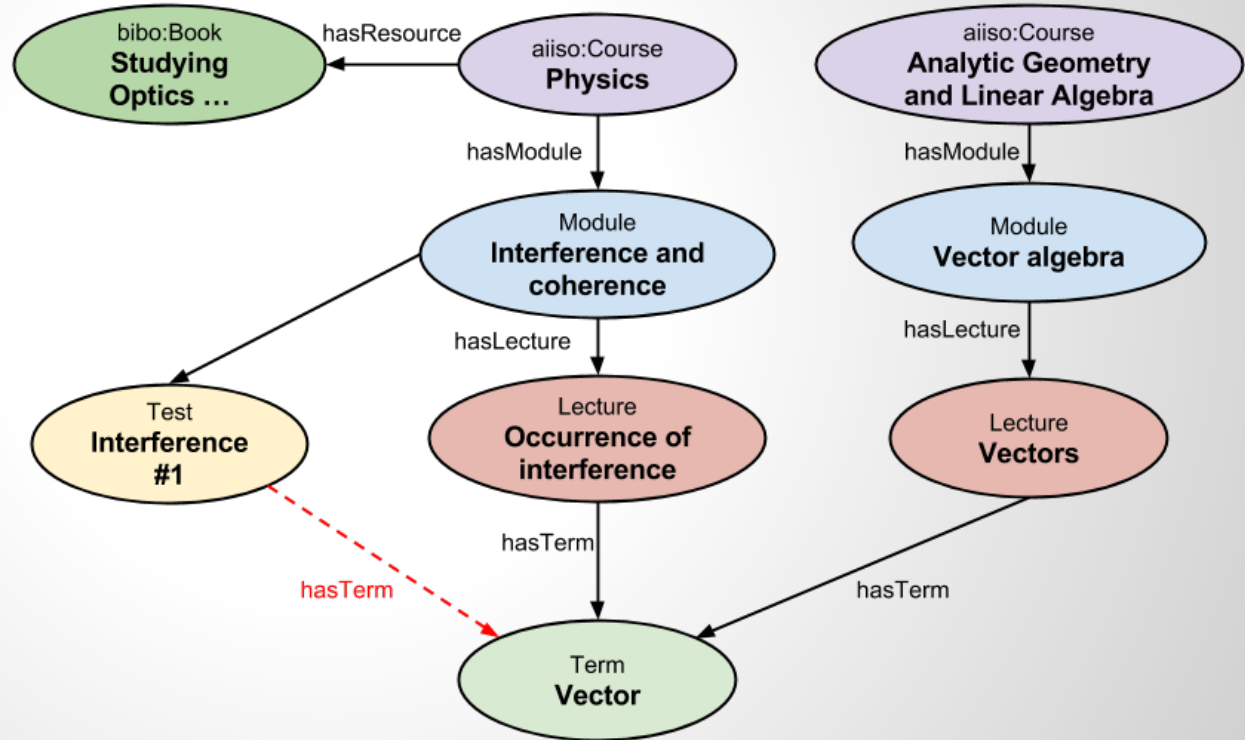
Linear algebra

For a general [vector space](#), the zero vector is the vector $\mathbf{0}$. The zero vector is unique: if \mathbf{a} and \mathbf{b} are zero vectors. The zero vector is a special case of the [zero tensor](#). It meaning the additive identity of the underlying [field](#). The [preimage](#) of the zero vector under a [linear transform](#) is the [kernel](#) of the transform. The zero vector is, by itself, [linearly dependent](#), and is the zero vector of any vector space. The zero vector is both parallel and perpendicular to any vector.



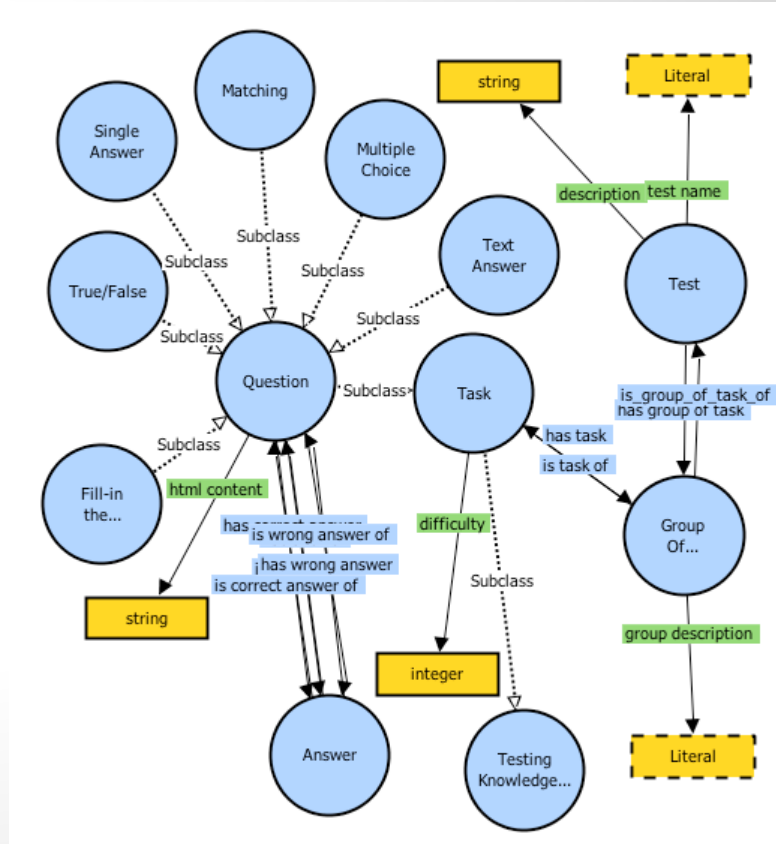
The ontology of education resources

- Extends AIISO ontology for education process and structure
- Uses BIBO for bibliographic resources
- Uses MA-ONT for media resources



The ontology of tests

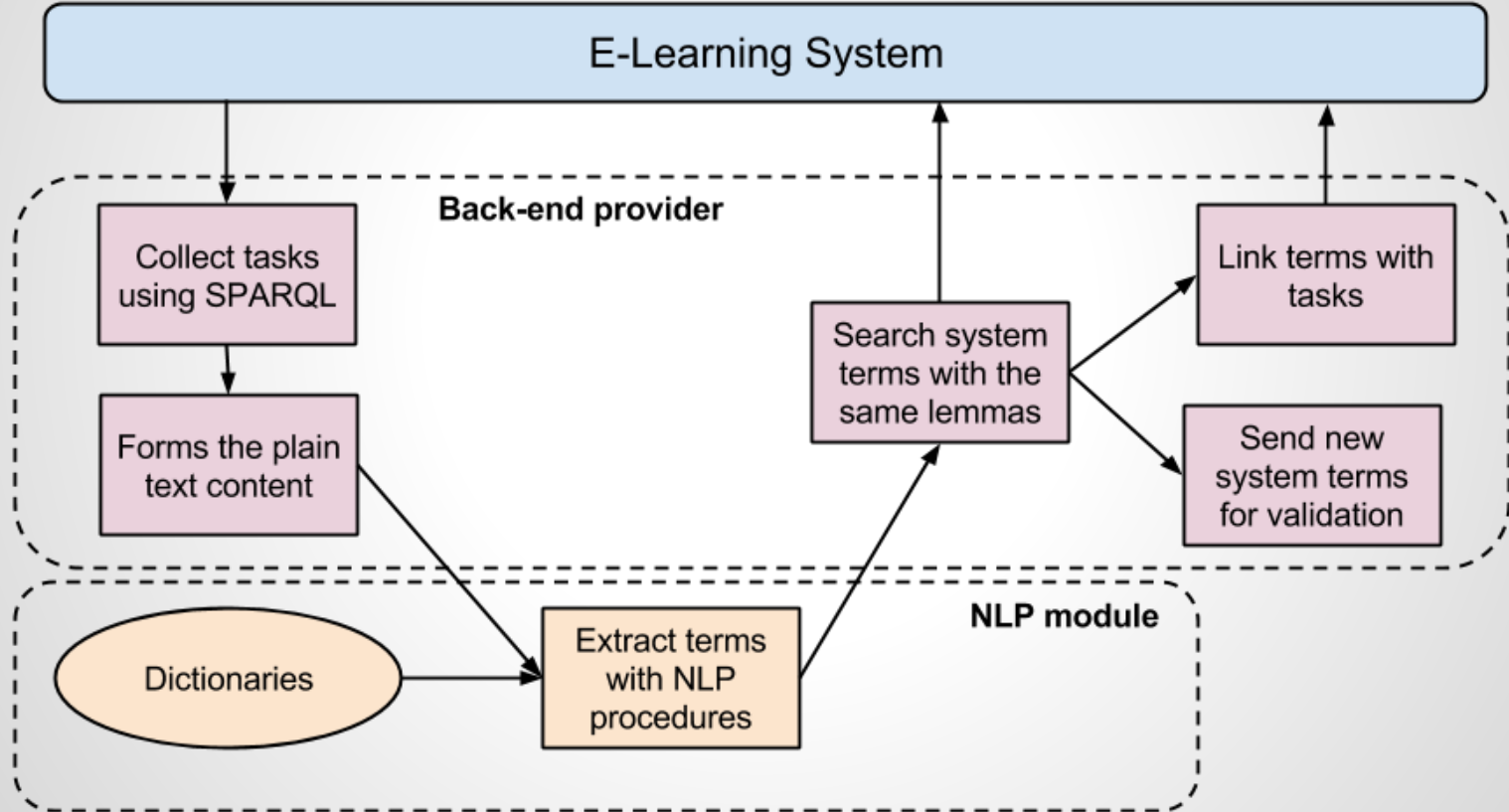
- Developed to describe the content of tests,
- Extends top-level ontology of the system,
- 12 classes,
- 10 object properties,
- 6 datatype properties,
- <http://purl.org/ailab/testontology>.



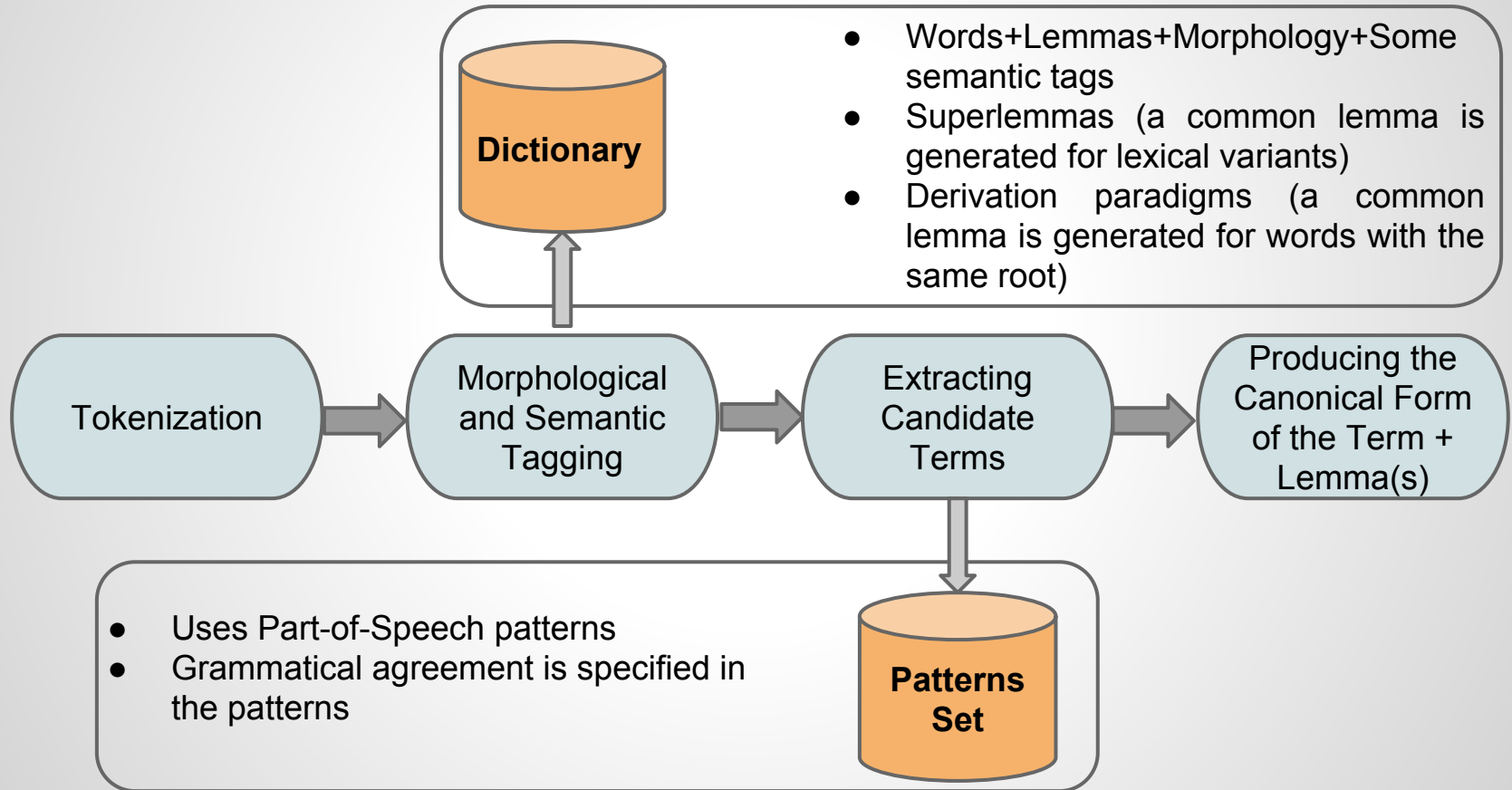
Lemmatization of Domain Terms

- Extracts lemmas from subject term labels,
- Uses NLP procedures and dictionaries to generate lemmas,
- Stores lemmas in triples.

Terms extraction and linking



NLP Algorithm



New System Terms

- If the extracted candidate term doesn't match any system terms, it is included as a new system term,
- New system terms are validated via SPARQL queries to DBpedia

```
SELECT DISTINCT ?term {  
  ?term dct:subject ?subject .  
  VALUES ?subject {  
    category:Concepts_in_physics  
    category:Physical_optics  
    category:Optics}  
  {?name_uri dbpedia-owl:wikiPageRedirects ?term ;  
   rdfs:label ?label .  
  }  
  UNION  
  { ?term rdfs:label ?label }  
  FILTER( STR(?label) = "Diffraction" )  
}
```



Implementation: Test parsers

- Converts tests of the course from XML to RDF,
- Uses Information Workbench XMLProvider to automatically update,
- Describes mapping using XPath functions.

The input XML code
<pre><test module="m_InterferenceAndCoherence" module_ns="Physics" uri="TestOfInterferenceAndDiffractionFrenel" name="Test Of Interference And Diffraction Frenel"> </test></pre>
The mapping code
<pre><rule id="test" nodeBase="//test" owlType="learningRu:Test" instanceNamespace="openeduTests" objectId="{./@uri}" objectLabel="{./@name}"> <objectPropertyMapping nodeBase="." instanceNamespace="openeduTests" value="{./@name}" owlProperty="ifmotest:hasGroupOfTasks" referredRule="task_group" /> </rule></pre>
The output RDF/XML code
<pre><rdf:Description rdf:about="http://openedu.ifmo.ru/tests/ TestOfInterferenceAndDiffractionFrenel"> <rdf:type rdf:resource="http://www.semanticweb.org/ k0shk/ontologies/2013/5/learning#Test"/> <label xmlns="http://www.w3.org/2000/01/rdf-schema#" Test Of Interference And Diffraction Frenel </label> <hasGroupOfTasks xmlns="http://www.semanticweb.org/ fedulity/ontologies/2014/4/untitled-ontology-13#" rdf:resource="http://openedu.ifmo.ru/tests/ Test_Of_Interference_And_Diffraction_Frenel"/> </rdf:Description></pre>

Implementation: NLP module

- Uses dictionaries in NooJ format:
<LEMMA>+<PART OF SPEECH TAG>+
<INFLECTIONAL PARADIGM>+<OTHER ANNOTATIONS>

air,N+FLX=TABLE

Michelson,N+ProperName

- English NooJ resources are reused, Russian lexical resources are original,
- A separate procedure implemented in Python is launched for lemmatization and term extraction.

Evaluation and Results

Percent of linked tasks, %	95
Percent of non-linked tasks, %	5
Number of different candidate terms	155
Number of manually extracted terms	30
Percent of system terms, matched by candidate terms, %	50
Percent of candidate terms, matched by system terms, %	8
Percent of candidate terms to be included to the system terms after the validation procedure, %	6
Percent of false candidates, %	86

Hidden subject terms

A ladder is 5m long. How far from the base of a wall should it be placed if it is to reach 4m up the wall?

Nothing to extract!!!

Hidden subject terms: Pythagorean Theorem, Hypotenuse, Cathetus.....

Conclusion

- The ontology of tests, dictionaries and patterns, test parsers, etc. have been developed,
- The tasks of the test have been linked with system terms,
- Statistics gathering module has been developed.

Future Work

- Improve term extraction procedure by adding parallel texts of tasks,
- Process units of measure in tasks to predict “hidden terms”,
- Use relations between subject terms to improve the quality of term extraction procedure,
- Refine term knowledge rating by replacing it by the proper ranking formula.

Thank you!!!

The front-end of the e-learning system: <http://ecole.ifmo.ru>

Example of subject terms analytics for module "Interference and Coherence":

http://openedu.ifmo.ru:8888/resource/Physics:m_InterferenceAndCoherence?analytic=1

The source code: <https://github.com/ailabitmo/linked-learning-solution>