

Implementing a Natural Language Processing Approach for an Online Exercise in Urban Design

Aleksei Romanov¹
romanov@corp.ifmo.ru

Dmitry Volchek¹
dvolchekspb@gmail.com

Artem Chirkin²
chirkin@arch.ethz.ch

Dmitry Mouromtsev¹
mouromtsev@mail.ifmo.ru

Arina Sender¹
arisend@gmail.com

Andrey Dergachev¹
dam600@gmail.com

¹ ITMO University
Saint- Petersburg, Russian Federation

² ETH Zürich
Zürich, Switzerland

Abstract

This paper presents the analysis of an online educational experiment. The idea of the experiment consists in using an interactive exercise platform to complement a Massive Open Online Course (MOOC) in Urban Design. The platform provides students an opportunity to map different urban structures and premises within a district or city, create notes and descriptions, so they can express themselves and share their views. At the same time, the platform uses students' responses to grade their submissions and records students' actions to aid the research in Urban Design. First, we overview the platform functionality and the exercise setting. Then, we take a closer look at the data sources provided by the platform. Finally, we describe the method that combines techniques of natural language processing and semantic technology to analyze students' responses and feedback. The analysis allows a better understanding of students ideas and concerns related to the problem the students are asked to solve.

Keywords: *online educational, Massive Open Online Course, urban design, natural language processing, semantic technology.*

1 Introduction

The object of the presented study is the natural language response of students of a Massive Open Online Course (MOOC). The MOOC used in this study is called “Future Cities”, is conducted by the chair of Information Architecture at ETH Zurich and ETH Future Cities Lab, and is hosted on edX site¹. The MOOC teaches the understanding of a city as a whole, its people, components and functions and is designed along the lines of Citizen Design Science. The main principles behind their approach are described by Mueller et al. in an article “Citizen Design Science: A strategy for crowd-creative urban design”[Mueller et al. 2018]. The online course consists of a series of videos, questionnaires, and exercises. The focus of the presented study is a pair of exercises in Urban Design provided by an external tool called Quick Urban Analysis Kit (qua-kit). The main problem of analyzing the results of the course study is a large number of text materials, such as reviews and discussions and the needs for their analysis. The article describes the method of the analysis that uses NLP algorithms. This methods allows us to extract the most common concepts being operated by the MOOC students describing their design submissions. The analysis of the concepts gives us an insight into the ideas of the students.

2 Qua-kit

Qua-kit is a web platform for simple urban design editing, sharing and discussion². Qua-kit is used in the MOOC in the form of two exercises:

1. *Design* exercise – work on a single design submission for a predefined scenario.
2. *Compare* exercise – given a series of twenty randomly selected submission pairs, select (vote for) one design in a pair that performs better according to a given design criterion.

Both, design and compare exercises require students to consider a list of design criteria. In qua-kit, a design criterion is presented by a description and an illustration; every exercise is linked with a list of one to four criteria.

Design exercise In the first qua-kit exercise a student is asked to redesign a part of container terminal area in Tanjong Pagar, Singapore. Figure 1 presents qua-kit user interface for this exercise. The tool uses WebGL to manipulate 3D geometry in browser. The student can move, delete, or create (from templates) individual objects. After the work is finished, the student submits the design with optional textual explanation of their ideas. At any moment, the student can come back to the site and update their submission. Other students can open the submission and write reviews (Figure 1, a panel on the right). A review consists of a criterion id (shown as an icon), a like/dislike tag, and an optional textual explanation. Therefore, the design exercise provides two sources of natural language feedback: the submission descriptions and the user reviews.

¹<https://www.edx.org/course/future-cities-ethx-fc-01x>

²<https://github.com/achirkin/qua-kit>

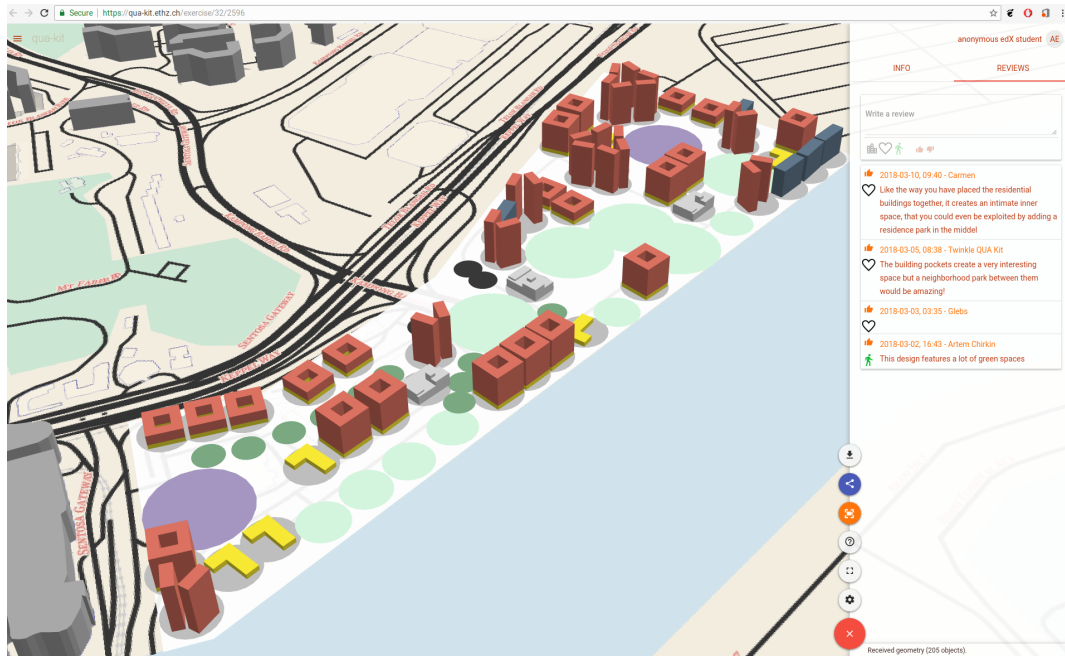


Figure 1: Qua-kit design exercise. A student designs their proposal in a 3D editor. Other students can review the design by giving comments and like/dislike tags (panel on the right).

Compare exercise In the second exercise, a student is asked to assess a series of design pairs. Figure 2 shows the interface for comparing two designs in a pair.

At the top of the page, the student sees the name and the icon of a design criterion under consideration. The student sees previews and descriptions of two randomly chosen designs. Then, the student has to select the one that seems to be better according to the given criterion by clicking on it. Optionally, a student can add a textual explanation for their choice.

Grading and the gallery When a number of students finish both of the exercises, qua-kit possesses enough statistical data to assign grades to the designs based on peer voting. Qua-kit runs a ranking system that updates design ratings according to the design criteria based on the votes from the compare exercise. Figure 3 shows the qua-kit gallery with student submissions. The gallery is available live at <https://qua-kit.ethz.ch>. The grades in qua-kit are assigned for each criterion independently. The algorithm of the ranking system may be interpreted as follows:

- Design the weighted majority of best *compare* voters defines the submission grade;
- Compare the voter gets the better grade if their votes agree with the majority.

Qua-kit constantly updates relevant grades when a user submits a design, updates a design, or votes for a pair of designs. Once a day, qua-kit averages the per-criterion-grades for an updated design submission or a voter and sends the result to edX. This way, the students get graded in the MOOC.

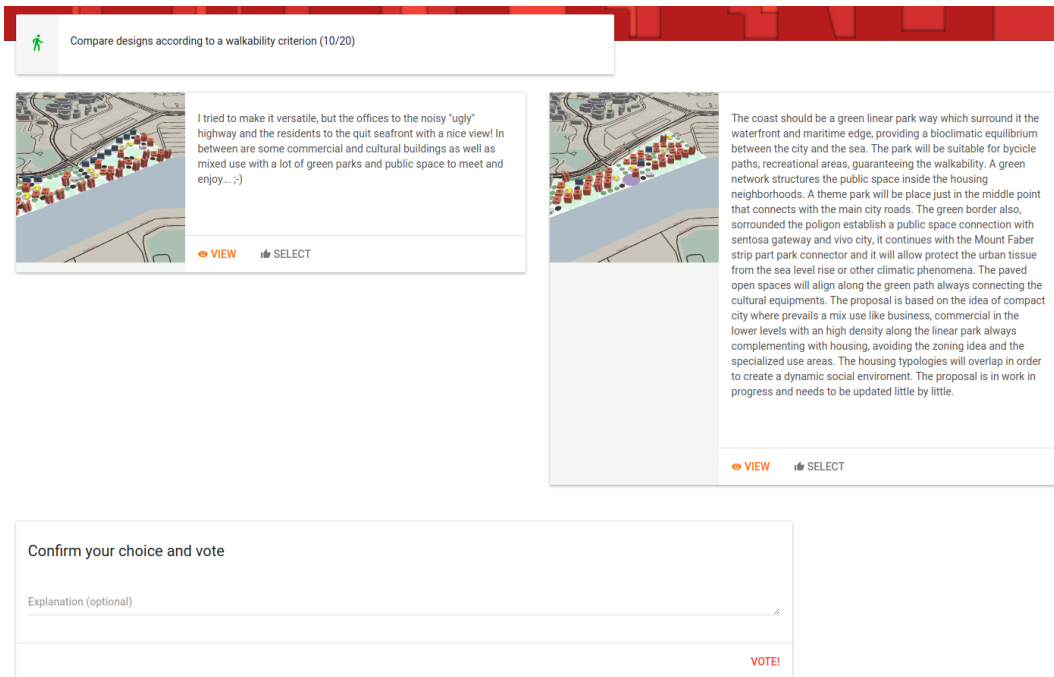


Figure 2: Qua-kit compare exercise. A student sees a series of design pairs and chooses which one is better according to a given criterion.

2.1 Collecting data

Qua-kit records all kinds of students activities: design (geometry) changes, new submissions, comments, reviews, and votes. This opens many possibilities for a research in Urban Design and education. In particular, the data allows studying a behavior of a student designer. Qua-kit designer interface as a tool defines the ways a designer can interact with a design [Volchek et al. 2017]. The simplicity of the tool enforces strong constraints on a designer expressiveness and limits the understanding of the design context. For example, by observing individual submissions we have found that some students have different understanding of what individual object models represent in reality, or different understanding of building density in the same design context. Therefore, one goal of the presented research is to study the relationship between a student submission and student opinions (description and reviews) on it.

To get an insight into student opinions and feedback in qua-kit, we need to analyze the natural language response of the students. The moderate amount of data produced by the exercises suggests for the use of natural language processing techniques (NLP). As follows from the Section 2, qua-kit provides three sources of textual data:

1. descriptions attached to the design submissions;
2. user reviews (like/dislike w.r.t. a criterion);
3. comments attached to the compare votes.

Unfortunately for the analysis, all types of textual feedback are optional. The most informative source of text is the submission descriptions. The students spend a lot of effort developing their submissions and usually feel eager to write at least a few sentences

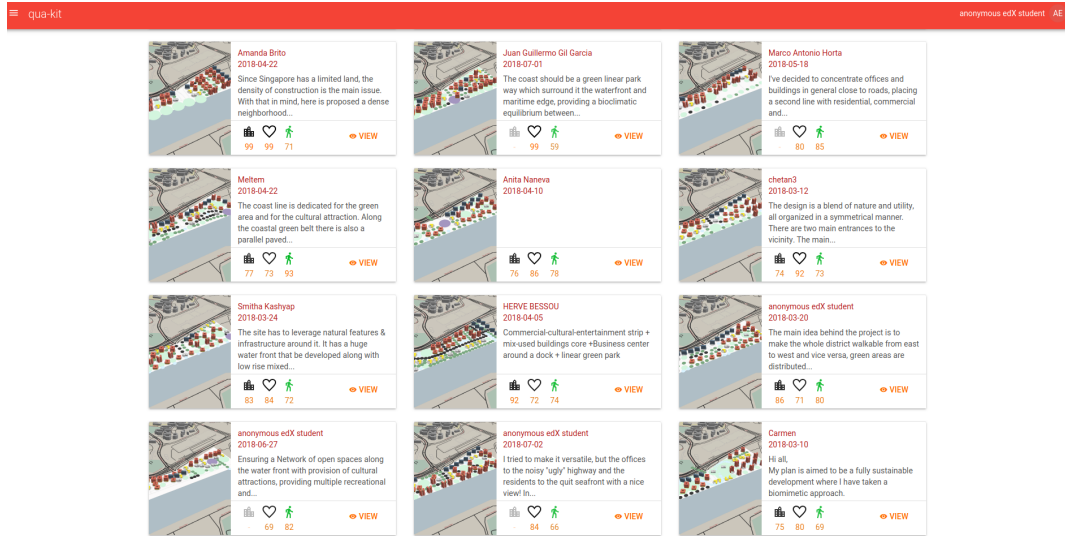


Figure 3: Qua-kit design gallery. Students submit their designs together with (optional) descriptions. These submissions are graded based on peer-to-peer evaluation and presented in the gallery.

about their ideas. The user reviews often contain comments, but the total number of reviews is not very large, because the reviews are not the part of the edX exercises. Lastly, the students rarely add comments to their compare votes, because this implies increasing the effort for the compare exercise significantly.

3 Method

Semantic technologies allow us to represent and analyze large volumes of information. This approach ensures the fulfillment of the task in a short time, assumes a high speed of operation and flexibility of the developed system. In addition, the use of the RDF description language³ allows us to present information in a form that is convenient not only for humans, but also for possible subsequent processing by machine methods (robots, crawlers, etc.). To achieve this goal, it is necessary to:

- **Extract existing data from the PostgreSQL database.** The analysis will require submissions sent by users, namely a text description of the completed task;
- **Extract the concepts using NLP algorithms.** In particular, it is necessary to use algorithms, which analyzes the presented text and returns a list of the concepts present in this text;
- **Develop an ontological model.** The model should represent the subject area in a semantic form, namely the main objects, the relationships between them and the attributes;
- **Map extracted data.** Only correct mapping will allow to execute queries and analyze results;

³<https://www.w3.org/RDF/>

- **Deploy the RDF-storage.** Tripplstore is necessary for storing all the received data, and also provides the interface for execution of queries (SPARQL-endpoint);
- **Define and create queries.** Queries are designed to extract the necessary information from the RDF-storage, for its subsequent interpretation and analysis. Also queries check the correctness of the mapping procedure. If the queries are not executed correctly - it is necessary to improve the mapping process;
- **Analyze the results.**

3.1 Data

The obtained data consists of user exercise submissions, their reviews, grades and votes in different criterion stored in PostgreSQL database. All data is easily obtained by executing queries. All data is processed in triplet format for import into RDF storage.

3.2 Concept extraction

The key thing in keyword extraction is language. The difference between nouns, verbs, adjectives and adverbs of a language is extremely useful in many natural language processing tasks. There are some fundamental techniques in NLP, including sequence labeling, n-gram models and evaluation.

POS-tagging is an important step for text analyzing. This step consists in the classification of words into their parts of speech also known as word classes or lexical categories and labeling them. A POS-tagger processes a sequence of words and attaches a part of speech tag to each word. For these tasks corpus readers are used. They provide a uniform interface for tagging.

For the purposes of extracting keywords, you should pay attention to such parts of speech as: nouns and nouns after determiners and adjectives[Kim Su Nam et al. 2013]. Adjectives and adverbs are important word classes too. Adjectives describe nouns and can be used as modifiers or in predicates. Adverbs modify verbs to specify the place, direction, etc. and may also modify adjectives. Most common approaches for tagging are regular expressions and unigrams. The regular expression tagger assigns tags to tokens using pattern matches. The unigram taggers use statistical algorithms.

But it is not enough to know the parts of speech, you need to work with the text itself. Tokenization is a way to split text into parts (tokens). These tokens could be sentences, or individual words. There are several approaches to divide the text. We used a verbose regular expression which consists of checking abbreviations, hyphens compound words, currency, percentages and different separate tokens.

In the processing of natural language is important from where you retrieved the data. Our data has been received from the reviews and descriptions of the student's home tasks. In such circumstances, it is necessary to consider possible typos and errors. So, grammar correction is necessary.

Final step is lemmatization words. It is a process of grouping together the different inflected forms of a word. The main goal is determining the lemma for a given word. It is the complex task, depending on the concepts that need to be obtained. Types of buildings, places, zones, roads, etc. play an important role in this project. We used complex metrics considering hyponyms of the word and semantic description. In addition, an important

role is played by the names of streets, famous buildings, districts, cities. Information on such ownership can be obtained from corpora (text corpuses). [Kovriguina et al. 2017]

All these methods allow to obtain concepts related to the field of Urban Design. As a result, it is possible to carry out a comprehensive analysis of the data of completed tasks by students.

The development consisted of the steps considered in the method. For the purposes of correcting spelling we used Python autocorrect library⁴ - simple implementation that everybody can use to implement a basic autocorrect features.

The NLTK library was used for POS-tagging. Tokenization is made by regular expression.

```
./.  
Creating/VBG  
(NP (NBAR privacy/NN))  
  inside/RB  
  and/CC  
  open/VB  
  to/TO  
(NP (NBAR new/JJ public/JJ spaces/NNS))  
./.
```

Tokens were selected according to the grammatical mask NBAR in sentences NP, with POS standard abbreviations: NN - Noun, singular or mass, NNS - Noun, plural, NNP - Proper noun, singular and JJ - Adjective.

```
NBAR:  
{<NN.*|JJ>*<NNS|NN>}  
{<NNP>*<NNP>}  
NP:  
{<NBAR>}  
{<NBAR><IN><NBAR>}
```

Thus, the concepts or phrases are obtained. First of all, concepts have been lemmatized by the WordNet lemmatizer, which take the POS tag into account. WordNet is a large lexical database of English. Nouns, verbs, adjectives and adverbs are grouped into sets of cognitive synsets. Synsets are interlinked by means of conceptual-semantic and lexical relations.

```
Concepts = ['privacy', 'new public space', 'community']
```

So, after tagging and lemmatization all concepts converted to synset format. This format allows us to access possible synonyms and hyponyms.

```
house  
  [Synset('building.n.01')]  
area  
  [Synset('region.n.03')]  
city  
  [Synset('municipality.n.01')]
```

⁴<https://pypi.org/project/autocorrect/>

In addition, based on the theme of the project, the concepts belonging to important geographical places and city spots were retrieved. This is possible due to semantic enrichment of concepts and named entity recognition (NER). We used machine learning model based on Groningen Meaning Bank (GMB) corpus⁵.

```
(S
  (org Tanjong_Pagar/NNP)
  is/VBZ
  a/DT
  historic/JJ
  district/NN
  located/VBN
  within/IN
  the/DT
  (org Central_Business_District/NNP)
  in/IN
  (geo Singapore/NNP))
```

As we can see 'Tanjong Pagar', 'Central Business District', 'Singapore' are members of the 'org' - organization and 'geo' - geographical groups.

Additionally all concepts were semantically searched among open data sets, such as Wikidata⁶, DBpedia⁷, etc. This allows concepts to be grouped together. As an example of the results of such a grouping by the "way" criterion.

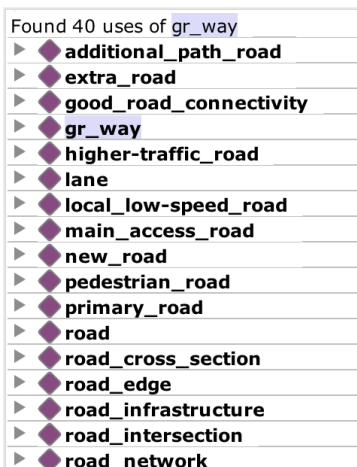


Figure 4: Concepts in group

3.3 Ontology development

Ontological modeling is one of the most important stages in works like this. The developed model should not only adequately reflect the subject area, its main classes and relations between them, but also contain links to a top-level ontologies.

⁵<http://gmb.let.rug.nl>

⁶<https://wikidata.org>

⁷<http://wiki.dbpedia.org>

Our ontology is based on the following top-level ontologies[Keßler et al. 2013]. Firstly, AIISO⁸ that provides classes and properties to describe the internal organizational structure of an academic institution. Secondly, FOAF⁹ (an acronym of Friend of a Friend), the ontology describing people, their activities and their relations to other people and objects. Finally, TEACH (Teaching Core Vocabulary)¹⁰, a lightweight vocabulary providing terms teachers use to relate objects in their courses.

We used Protégé¹¹ for ontology development and Ontodia¹² for visualization. A part of the ontology that describes the relations between concepts, submissions, and reviews is shown in the figure 5.

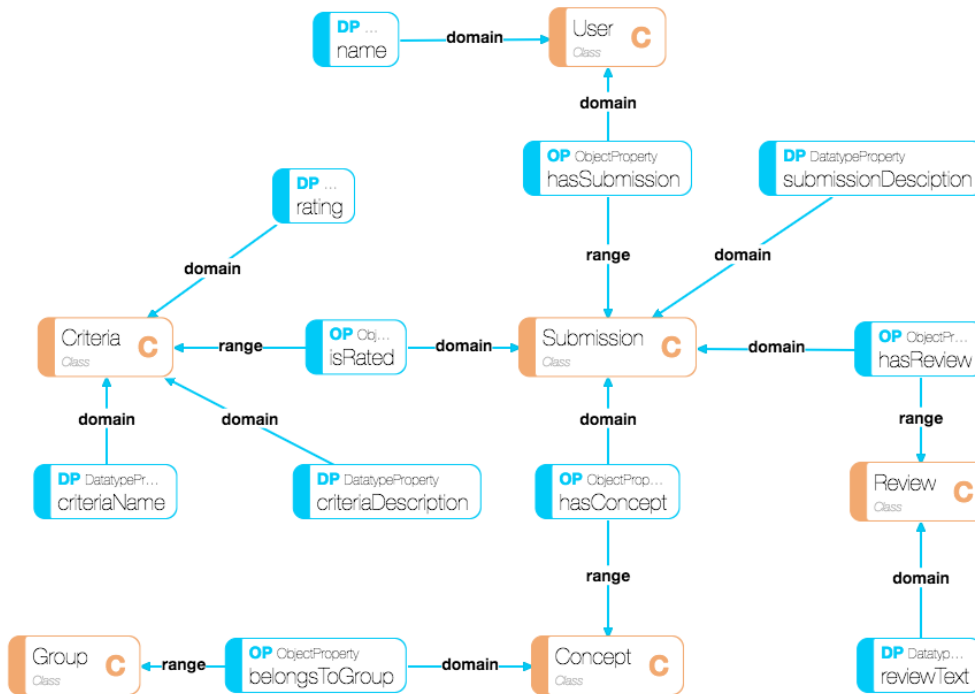


Figure 5: Submission and Concept ontology

3.4 Mapping

Since we have the ontology and the extracted concepts it becomes possible to start mapping procedure. We used N-triples format to form RDF-storage. To map the concept on the ontology it's necessary to make several steps: to define that the concept is a "Named Individual", it belongs to class "Concept" and it may belong to a "Group". The example is shown below.

```
Urban:SomeConcept rdf:type owl:NamedIndividual.
Urban:SomeConcept rdf:type Urban:Concept.
Urban:SomeConcept Urban:belongsToGroup Urban:someGroup
```

⁸ <http://purl.org/vocab/aiiso/schema>.

⁹ <http://xmlns.com/foaf/spec/>

¹⁰ <http://linkedscience.org/teach/ns/teach.rdf>

¹¹ <http://protege.stanford.edu>

¹² <http://ontodia.org>

Mapping of the rest of the extracted data on the ontology was done in a similar way.

4 Results

To evaluate the results of the project, it is necessary to analyze the data, namely, to identify the main dependencies, to estimate the existing trends. The resulting triple store contains 36,926 statements. To extract the necessary information, the query language SPARQL was used. For example, a query to display the frequencies of mentioning concepts is presented below:

```

PREFIX Urban:
<http://www.semanticweb.org/Urban>
select ?concept (COUNT(?concept) AS ?all)
where ?submission Urban:hasConcept
?concept.
group by ?concept
order by desc(?all)

```

Table 1: Rates according the criteria

N	Criterion	Rates
1	Visibility	670
2	Centrality	619
3	Connectivity	496
4	Accessibility	470
5	Accesibility/Connectivity	225
6	Density/Distribution	203
7	Walkability	11
8	Personal preference	10

The distribution is presented in the table 1. 1491 user participated in the exercise, however there is only 721 tasks submitted. Peer assessment was used to evaluate the results of the exercise. A total of 308 reviews were left. At the same time, only 88 submissions had at least one review. Submissions were rated according to the offered criteria. A total of 2704 rates were left.

To analyze the relation of what students say and what they actually implement we extracted the concepts from the submissions and the reviews. A total of 1816 concepts were extracted. The distribution of the most popular concepts is presented in the figure 6.

A significant part of the most popular concepts consists of too common ideas like "building", "road", "city" which have neutral meaning and do not reflect user preferences. If we exclude such a concepts, we can analyze concepts that really reflect user's vision.

It is possible to group concepts into several main clusters and find out the popularity of each cluster. The results are presented on the fig.7 Among them:

- **People.** This cluster includes concepts 'people', 'community', 'neighborhood', etc and describes the care about a person and her relationships with the community.

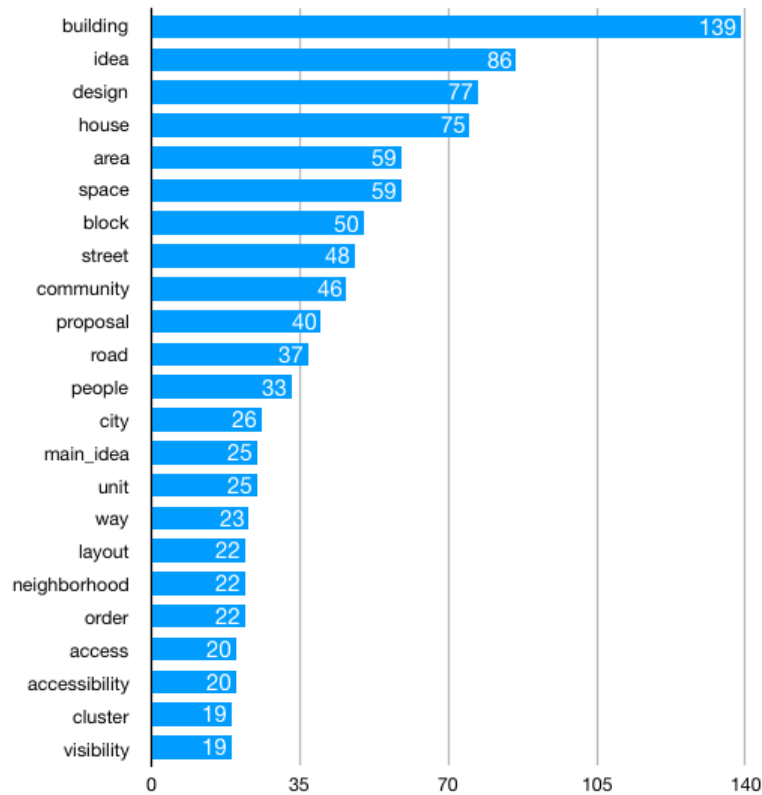


Figure 6: Number of mentions of the most popular concepts

- **Space.** Includes concepts, responsive for the description of the environment 'space', 'open space', 'public space', etc.
- **Centrality.** The criterion is used to understand structural properties of complex relational networks. Centrality measures identify that, in a network, some nodes are more central than others.
- **Green areas.** Includes concepts like 'park', 'green area', 'green space' etc.
- **Visibility.** Visibility analysis shows the visual impact from a point into the surrounding environment affected by the obstructions and shaping the skyline. In the city, the urban elements such as topography, building, trees, etc., make part of the urban atmospheric visibility.
- **Accessibility.** Accessibility can be defined as the facility in which one place can be reached from another place and is dependent on the spatial distributions the given location.
- **Water.** This cluster includes concepts that describe water environment: 'water', 'waterfront', etc.
- **Connectivity.** For urban networks, connectivity is used to understand spatial conditions affecting pedestrian activities and behaviors in cities.

- **Density.** Density can be defined as the mass of an object per unit area. In architecture and urban planning, physical density is a numerical measure of the concentration of individuals or physical structures within a given unit area.

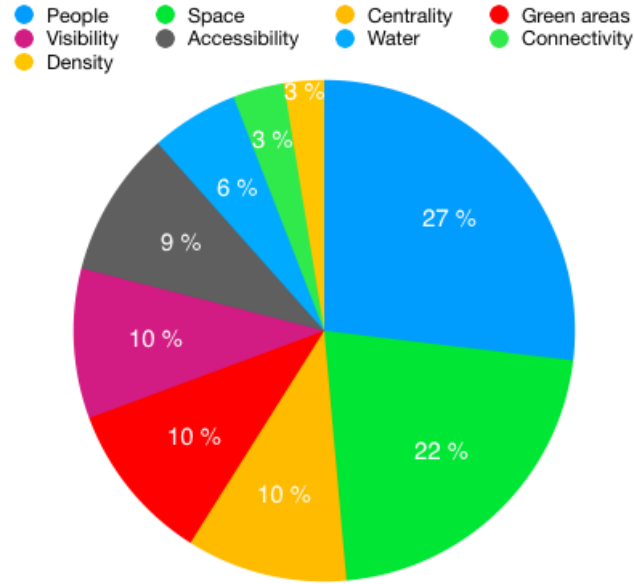


Figure 7: Cluster popularity

Adjectives are used to describe and clarify the types of zones, places, etc. The list of the most popular and significant adjectives, encountered in the submissions is presented on the figure 8

According to this information students were highly concerned about caring for a person, her needs, and her representation as a member of society. Second important aspect is space organization. Centrality, Visibility, Accessibility, Connectivity and Density are the criteria by which the works were rated, so students are expected to mention these terms in their submissions. Finally parks, green zones and water environment were quite popular topics in the submissions.

Students were suggested to do several tasks and these tasks were assumed to be rated by the different criteria. We united similar criteria and counted the number of the rates and the number of the corresponding concepts mentions in the submissions. Results can be seen in the table below:

Table 2: Rates according the criteria

Number	Criterion	Rates	Mentions
1	Accessibility	695	40
2	Visibility	670	41
3	Centrality	619	44
4	Connectivity	496	14
5	Density/Distribution	203	11

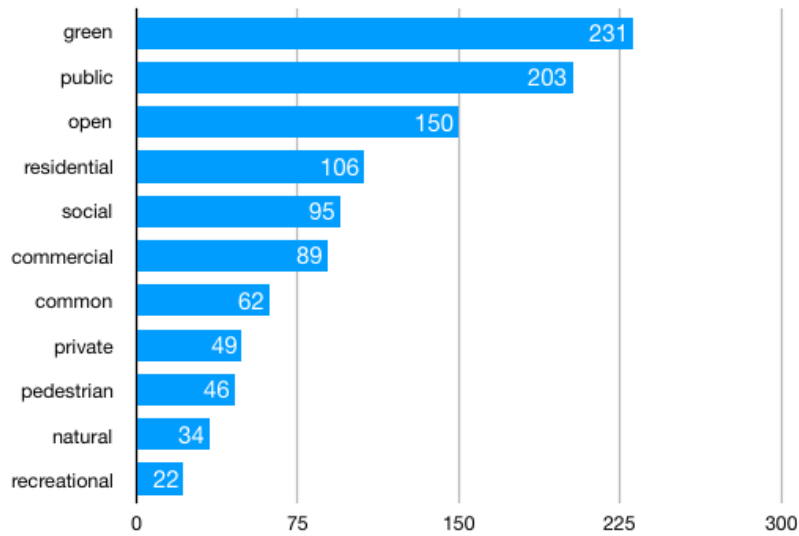


Figure 8: Adjective popularity

According to the data among 721 submissions, concepts reflecting the criteria, are mentioned only in 30 submissions on the average. But there is a connection between number of Rates of each criterion and the number of corresponding concepts. The Pearson correlation coefficient is sufficiently high (0.86). Thus it's possible to say that if the submission contains the concept, reflecting the corresponding criterion, it has more chances to be rated.

The above data collection is the result of processing all tasks performed by students. Finally, the task from Responsive Cities [EdX FC-04x-2] was considered separately.

We used LDA for the purposes of data visualization. LDA is an unsupervised technique, meaning that we don't know prior to running the model how many concepts exists in our corpus. Topic coherence, is one of the main techniques used to deestimate the number of topics [Rosner et al. 2018] [Muhammad Omar et al. 2015]

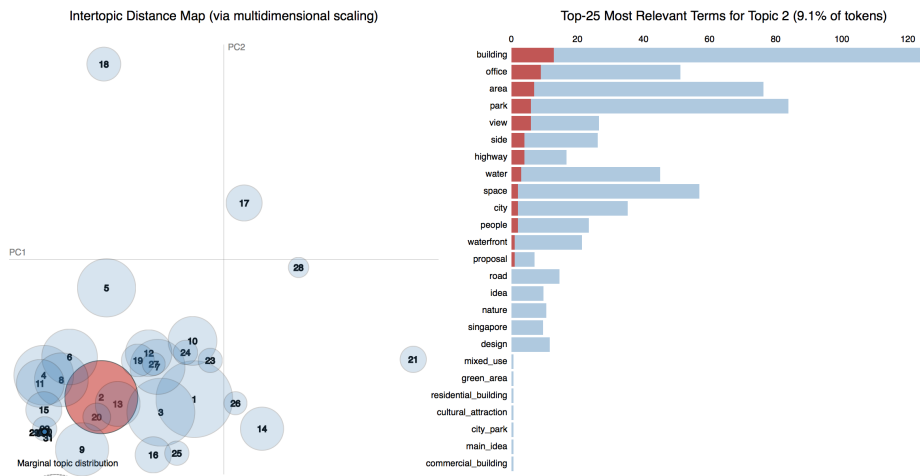


Figure 9: Adjective popularity

On the left side of figure 9 we can see circles representing different topics and the

distance between them. This approach allows us to clearly see the density of the tasks descriptions, as well as to identify any anomalies.

5 Conclusion

Qua-kit was used for conducting MOOC exercises. It provided the data of peer reviews and task descriptions for subsequent analysis. We proposed the method to identify and classify keywords reflecting the concepts operated by MOOC students. Keywords and exercise materials such as peer reviews and descriptions were represented as the RDF storage. It allowed handling these materials by sending queries that are natural for humans. Then, we performed the LDA analysis of the data. We analyzed how the students described their design submissions and evaluated the submissions of their peers.

In future, we plan to integrate and automate the method to improve tasks. Additionally, we are looking to automating verification of peer assessments and search for anomalies. Another research direction is combining the analysis of submission content (geometry) and the analysis of submission descriptions to study the relations between the two.

Acknowledgement

This work is partially supported by Scientific Technological Cooperation Program Switzerland-Russia (STCPSR) 2015 (project IZLRZ1164056).

The work is partially supported by the RGNF grant 16-23-41007.

References

- [Keßler et al. 2013] Keßler, Carsten and d’Aquin, Mathieu and Dietze, Stefan (2013) Linked Data for science and education. // *Semantic Web*, p. 1–25.
- [Kim Su Nam et al. 2013] Kim, Su Nam and Medelyan, Olena and Kan, Min-Yen and Baldwin, Timothy (2013) Automatic keyphrase extraction from scientific articles. // *Language Resources and Evaluation*, p. 723–742.
- [Mueller et al. 2018] Johannes Mueller and Hangxin Lu and Artem Chirkin and Bernhard Klein and Gerhard Schmitt (2018) Citizen Design Science: A strategy for crowd-creative urban design. // *Cities*, p. 181 – 188.
- [Rosner et al. 2018] Rosner, Frank and Hinneburg, Alexander and Röder, Michael and Nettling, Martin and Both, Andreas (2013) Evaluating topic coherence measures. // *Conference: Neural Information Processing Systems Foundation (NIPS 2013)*, p. 23 – 47.
- [Kovriguina et al. 2017] Kovriguina, L., Shilin, I., Putintseva, A., Shipilo, A. (2017) Russian Tagging and Dependency Parsing Models for Stanford CoreNLP Natural Language Toolkit. // *International Conference on Knowledge Engineering and the Semantic Web*, p.101–111. – Springer, 2017.

- [Muhammad Omar et al. 2015] LDA topics: Representation and evaluation (2015) Russian Tagging and Dependency Parsing Models for Stanford CoreNLP Natural Language Toolkit. // *Journal of Information Science*, p.662–675.
- [Volchek et al. 2017] Volchek D., Romanov A., Mouromtsev D. (2017) Towards the Semantic MOOC: Extracting, Enriching and Interlinking E-Learning Data in Open edX Platform. // *Knowledge Engineering and Semantic Web. KESW 2017. Communications in Computer and Information Science*, p.662–675. – Springer, 2017.