

HAIKU POETRY GENERATION USING INTERACTIVE EVOLUTION VS. POEM MODELS

Miroslava HREŠKOVÁ, Kristína MACHOVÁ

Department of Cybernetics and Artificial Intelligence, Faculty of Electrical Engineering and Informatics,
Technical University of Košice, Letná 9, 042 00 Košice, Slovak Republic,
E-mail: miroslava.hroncova@student.tuke.sk, kristina.machova@tuke.sk

ABSTRACT

The article describes and compares two different approaches to haiku poetry generation. First approach uses interactive evolution to create poems and the second composes poems from haiku models. The goal is to create poems, considered by humans as understandable and with aesthetic value. Two systems were created, one implementing each approach.

Keywords: computational creativity, haiku, haiku poetry, human-robot interaction, natural language generation, poetry generation

1. INTRODUCTION

Natural language generation aims to develop computer systems that can produce understandable texts in human language. Since poetry is a specific genre of human language, task of poetry generation is a highly specialized instance of natural language generation.

The goal of this article is to propose, describe and compare two different approaches for haiku poetry creation.

Haiku poetry originated in Japan. It has fixed form – always 3 verses and each verse has given syllable count (syllable pattern 5-7-5). It is often inspired by nature and captures a feeling.

Generating haiku poetry was chosen because of its fixed form and simple idea of the content.

These approaches are implemented as web-based applications and generate haiku poems in English language. Performance of systems, that are the implementation of proposed approaches, was evaluated according following criteria:

- Form – the poem forms 3 lines and each line conforms to syllable pattern typical for haiku poems
- Content – generated poem contains words related to nature and/or expressing emotions
- Meaning – the poem can be considered as meaningful, there is idea recognizable behind the verses

In the following part of the article, description of several related works can be found. Section 3 and 4 is dedicated to description of the proposed approaches for haiku poetry generation, their implementation and overall evaluation, as well as example outputs. One possible poem meaning is added below each poem by author of this article. Since understanding and evaluation of poem is very subjective, it can differ from person to person and does not have to be the same as the one stated in this article. Last part of the article is dedicated to comparison of both proposed approaches and description of further work.

2. RELATED WORKS

Different systems use different approaches to generate natural language and poetry. Some of them are described in this section.

2.1. Generating Textual Summaries from Data

Several systems have been built that produce summaries in text form from large amount of data. For example, daily stock market reports generated from a stock quotes database or daily weather reports from a meteorological data.

System called Ana [1] uses data from a Dow Jones stock quotes database as input and produces opening paragraphs of a stock market summary as output.

2.2. Generating Weather Forecasts

The first system for generating weather forecasts by computer is FoG [2]. It was used in Canada to generate weather forecasts in French and English in the early 1990s.

A research in this area includes an experiment which showed that users sometimes preferred computer-generated weather forecasts to human-written ones. The main reason is that the computer-generated forecasts used more consistent terminology.

2.3. Generating Wikipedia Articles

Approach proposed in [3] is used to create an overview of a user-chosen subject. This overview is created as text from information drawn from the Internet.

Human-authored documents related to user-chosen subject are analysed and searched for patterns. This way, content template is created. Then, the system produces a new article by selecting content from the Internet for each part of this template. It can be used for automatic generation of articles on Wikipedia portal.

2.4. Generating Narratives

Tale-Spin [4] is a system creating stories by using algorithm. At first, it creates something like story world

that can be considered as simulation of a problem. The story world contains simulation of characters, their goals, personalities and their relations among each other. Stories are made by computing consequences of the model – accomplishing the goals of the characters in the story world.

Brutus [5] is another example of a program that composes narratives. At first, theme of the story is established by creating characters, places, goals and actions. Then, depending on the selected theme, the system develops a plot by planning and simulation. The structure of the story is created based on story grammars that contain dictionary and rules for creating sentences.

2.5. Eureka

Historical computer [6] called Eureka generated verses in Latin. It was created by John Clark in 1845. By pulling a lever, machine printed out a Latin verse, both grammatically and metrically correct.

Eureka was writing in hexameter - each line has six metrical units called feet.

To ensure that the verse made sense, words were grouped by similar meaning and the verses have identical syntactical order (word pattern – always the same: ADJECTIVE NOUN ADVERB VERB NOUN ADJECTIVE).

With a grant from the UK's Arts and Humanities Research Council, however, Alfred Gillett Trust, organisation that aims to preserve the history of the Clark family, restored the Eureka to working order. It was exhibited for a short time and now a full-scale modern replica of it is being created.

2.6. NSA Haiku Generator

Random generator of haiku poems [7] composes poems using NSA terms. It is a web application created to entertain. Poems are not conforming to haiku syllable pattern and seem to be random clusters of words.

2.7. Random Poem Generator

At [8], a poem generator is introduced where user is able to choose poem genre and poem theme. Following genres, defining the poem structure, are available: tanka, lanturne, nonet, quinzaine, haiku, katauta, sedoka and clarity pyramid. Provided themes consist of the elephant and albatross, the cloud and Pennsylvania '99. Also there is a possibility to select no theme. That means to generate random haiku without any topic restriction.

By clicking on a Generate button, poem is generated. The generation process consists of random verse selection from the verse corpus. The verse corpus contains large amount of poetry lines with different syllable count.

Poems are sticking to syllable pattern typical for haiku poetry. However, user has to generate several poems until being able to generate one that has meaning.

2.8. ThinkZone Poem Generator

Poem generator at [9] allows users to generate poems by filling pre-defined sentence patterns with words from word list.

Some sample sets of words are provided for user to choose from. They are thematically grouped into sea- and city-related. Words in sample lists are divided into categories based on their part of speech. Website also offers possibility to create custom word set. User can define 9 categories and fill them with own words.

Sentence patterns define sequence of word categories. For each category in sequence, random word from word list is selected.

To make a poem, random number of randomly selected sentence patterns is chosen and these are filled with words from the user-chosen word set.

The website offers possibility to use pre-defined word sets and also sentence patterns. Another useful feature is the possibility to customize (for more advanced users) them or create new from scratch.

The application does not take into account rhythmic side of poetry and does not implement counting syllables. Main disadvantage is that the web application only randomly selects words into pattern based on category and does not guarantee meaningfulness.

2.9. Haiku with Sounds and Images

System for haiku poetry generation described in [10] creates haiku poems. It is able to vocally interpret them using program for Japanese speech generation.

The input to the generator is blog website. It is used for determining haiku theme by searching for the most frequent nouns. Also adjectives are extracted from the input blog site.

To properly structure the poems, algorithm utilizes grammar templates from existing haiku poems. Available poem database contains haikus from haiku-masters and also amateur poets. The algorithm incorporates the use of kigo. The most suitable seasonal word is chosen related to haiku topic from a kigo database.

Also there is added value in sound effect and picture accompanying the poem. Sound is chosen from pre-prepared small sound set and image is searched on the Internet. Experiments showed that haikus accompanied with sound effects and thematic picture were evaluated higher.

The generated outputs are good, they are formally correct and also have meaning. Big plus is the implementation of using seasonal words.

3. HAIKU POETRY GENERATION WITH INTERACTIVE EVOLUTION

This approach for haiku poetry generation uses evolutionary algorithm to create new poem and leverages human evaluation as fitness function.

The approach was implemented as web application and it was also used with NAO robot and extended to evolution of recitation as a whole (includes evolution of poem, voice pitch and speed, gestures of robot) [11].

This article refers only to web application of the approach described in [11] where users are able to evolve poems personalized to themselves. Screenshot of evaluating screen is shown in Fig. 1.

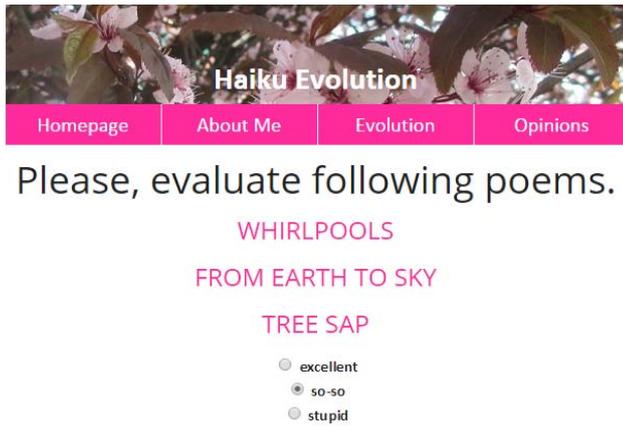


Fig. 1 Web application for haiku poetry generation with interactive evolution

3.1. Haiku corpus

The corpus is a database consisting of haiku poems created by human authors. It is used as the source for creating initial population and injection of new individuals in each generation.

Haiku poems from following haiku portals were used to create the haiku corpus:

- AhaPoetry [12]
- DailyHaiku [13]

3.2. Structure of individual

One haiku poem represents one individual. Haiku verse is considered the basic element of the poem. Each individual is structured as a collection of 3 verses.

majestic maples (1st verse)
because they are admired most (2nd verse)
when their leaves are old (3rd verse)

Fig. 2 Individual- haiku poem consisting of 3 verses

Therefore, chromosome of one individual has fixed size of 3 attributes. Attribute value is 1 haiku verse.

3.3. Genetic operators

Custom genetic operators were designed to fit the purposes of the haiku evolution. Since the basic element of poem, in this case, is verse, genetic operators are working with verses.

Two different genetic operators were designed:

- cross-over with two parents
- cross-over with three parents

Cross-over with two parents consists of exchanging a verse in a haiku at randomly chosen position with respective verse from another haiku. First parent comes from current generation and the second is selected randomly from haiku corpus.

spring wind
an old lady tries again
to tidy her hair

Fig. 3 1st parent

one ear
listening to the river
wet with rain

Fig. 4 2nd parent

Verse number selected for exchanging in 2-parent cross-over: 3

spring wind
an old lady tries again
wet with rain

Fig. 5 New individual created by crossing-over the selected 2 parents

For cross-over with three parents, first parent comes from current generation and two other parents are selected randomly from haiku corpus and new haiku is constructed from the three haiku poems in following way: it takes first verse from the first parent, second verse from second parent and the third verse from the third parent.

aging beauty
truth lies on the mirror
reversed

Fig. 6 1st parent

the dam is broken
spring cascades into valleys
as apple blossoms

Fig. 7 2nd parent

pears
coloring themselves
the days of summer

Fig. 8 3rd parent

aging beauty

spring cascades into valleys

as apple blossoms

Fig. 9 New individual created by crossing-over the selected 3 parents

3.4. Initial population and its evolution process

Each population contains 10 individuals. Population size was chosen experimentally – several experiments with real users were carried out and their opinion was taken into account when choosing the size of population for evolution.

Roulette method was chosen for selecting parents for reproduction. 8 parents are selected and each parent is then crossed-over by one randomly chosen cross-over operator.

Reproduction creates new population of 8 individuals, other 2 are selected randomly from haiku database.

3.5. Poem evaluation

At first, every poem in population has the same fitness value. Later in evolutionary cycle, the fitness value is modified based on human evaluation.

Evaluation possibilities are 3:

- Excellent – fitness value is doubled and the individual has higher probability to be selected for reproduction
- So-so – individual's fitness is not changed, default (non-zero) value is kept
- Stupid – fitness of the individual is set to 0 and therefore the individual cannot be selected to reproduce

3.6. Implementation of the proposed approach

Generation of haiku poems with interactive evolutionary computation was chosen because the user's subjective preference is the most important task when generating natural language, especially poetry. Goal state is a poem that satisfies the preferences of human.

Approach of interactive evolutionary computation is suitable also because of big search space of possible haiku poems. The proposed approach was implemented as web application.

3.7. System evaluation

The application performance depends very much on the haiku corpus and poems selected into initial population.

Evaluation criteria:

- Formal criteria – 100 % of output poems have 3 verses, however only about 45 % of generated poems conform to 5-7-5 syllable pattern. The reason is that not every poem in haiku corpus conforms to this pattern. With different corpus (from other sources) that would contain haiku

with proper form, the formal criteria would be met.

- Content – 100 % of generated poems contains vocabulary connected to emotion and/or nature.
- Meaning – approximately 40 % generated poems can be considered as meaningful.

3.8. Example outputs

More examples are provided in this section.

a sunny porch

shaded with an umbrella

quiet thoughts

Fig. 10 Haiku example 1 generated by interactive evolution

This haiku creates picture of porch that author considers perfect place for contemplating.

her tinkling laugh

weaving winter sunshine

with morning frost

Fig. 11 Haiku example 2 generated by interactive evolution

The poem listed above aims to capture author's impression of a laughing woman. The laugh reminds the author of "weaving the sunshine with frost".

at sunny porch

a dry leaf chases me home

suddenly silent

Fig. 12 Haiku example 3 generated by interactive evolution

This poem tries to create an image of sunny porch welcoming the author when he is returning home.

Often, the system creates less meaningful poems and poems with confusing content.

her anger

a loud dull swish

drying dill

Fig. 13 Haiku example 4 generated by interactive evolution

4. HAIKU POETRY GENERATION WITH POEM MODELS

Proposed approach to haiku generation constructs poems from words based on language models that take

into account the word's part of speech and its number of syllables.

The approach was also implemented as web application. Users are able to generate a poem and evaluate it. Generated poem is saved to database with its evaluation. This information is intended for further use to improve the approach when enough evaluations from real users will be gathered.

Screenshot of evaluating screen is shown in Fig. 14.

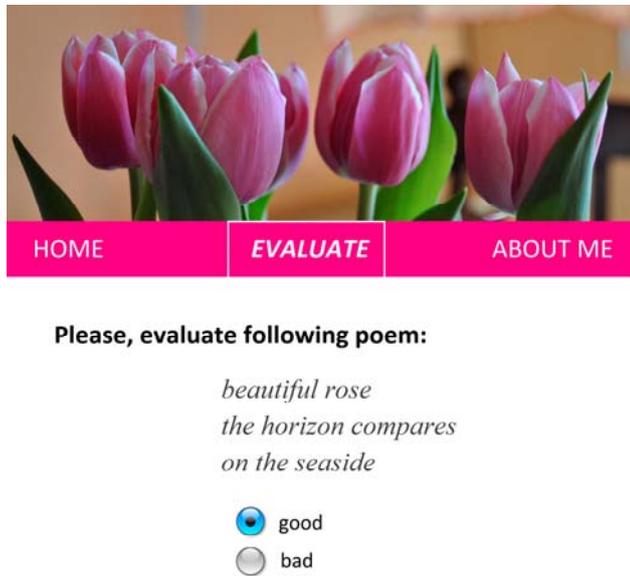


Fig. 14 Web application for haiku poetry generation with poem models

4.1. Haiku corpus

The corpus for both approaches is the same, but both systems use it for different purposes. In this case, it is the source for gathering the haiku specific words and extracting poem models.

4.2. Dictionary creation

Data preparation phase includes creating a word dictionary. It consists of haiku specific words, that were acquired from haiku corpus, and basic vocabulary consisting of general words.

This approach was selected to make sure that the vocabulary, used in generated poems, will be consistent with the general haiku content rules.

At first, set of haiku specific words is extracted from haiku corpus. Then, each word is pre-processed. This part consists of following steps:

- word is put to lowercase
- numbers and other non-alphabetic characters are removed
- unnecessary spaces are removed

When acquiring haiku-specific words, at first all stop words have to be removed. Stop words are very common words that have little value in text mining and are excluded from the vocabulary entirely. The main reason for removing stop words is that words that are used the most (e. g. the, a, be, in, ...) are words belonging to general vocabulary and therefore are not haiku-specific.

For every word in dictionary, metadata is determined. Metadata for each word consists of:

- part of speech
- syllable count
- number of occurrences- to acquire only haiku-specific words, only words with occurrence number higher than 10 are kept (occurrence number was chosen experimentally, it depends on the size of dictionary - the bigger dictionary, the bigger constant).

For creating meaningful poems, also general vocabulary is needed. Therefore, dictionary of common words is loaded from [14].

Metadata for each word belonging to general vocabulary is acquired (in the same way as for haiku-specific vocabulary, except of occurrence number).

All words from dictionaries are saved to database.

4.3. Model extraction

To make sure that proper haiku form is kept, models for haiku creation are generated from haiku corpus. A haiku model has form of a list of parts of speeches.

Then, occurrence number of each unique model is counted, because only models with occurrence number higher than 5 are kept as most used.

The occurrence number was chosen experimentally and depends on the size of haiku corpus (the bigger corpus, the bigger constant).

Examples of haiku model extracted from corpus:

*PREPOSITION ADJECTIVE NOUN
ADJECTIVE NOUN
ADVERB VERB*

Fig. 15 Example model

4.4. Syllable division

Random model from extracted model list is used. Syllables are divided randomly for each part of speech in the selected model based on strict form of haiku poem (5-7-5 syllable pattern).

Example of syllable division for example haiku models shown in previous subsection:

*1-2-2
4-3
3-2*

Fig. 16 Example syllable division (for example model in Fig. 15)

4.5. Haiku generation

Words are selected from dictionary based on selected part of speech and word's syllable count.

Example of haiku poem with example syllable division and example haiku model shown in previous subsections:

in scary garden
independent sunflower
quietly cowers

Fig. 17 Haiku poem generated based on example model in Fig. 15 and syllable division in Fig. 16

Generated haiku is displayed and saved to a database.

4.6. Implementation of the proposed approach

Approach for haiku poetry generation with poem models, described in this article, is implemented as web-based application.

The user interacts with its graphical user interface. The goal of the user is to evaluate haiku poems, generated by the system, in means of aesthetic impression and meaning.

4.7. System evaluation

The application performance depends very much on the dictionary and haiku models used for generating poems. That is why a large haiku corpus was used- to create "haikuness" in each poem.

Evaluation criteria:

- Formal criteria - 100 % of output poems have 3 verses with 5-7-5 syllable pattern and thus comply with haiku formal rules.
- Content - 90 % of all generated poems contains vocabulary connected to emotion and/or nature.
- Meaning - with general vocabulary from [14] and haiku specific dictionary and models extracted from [12] and [13], 60 % of all outputs can be considered as meaningful.

4.8. Example outputs

In this section, several more example poems are listed.

silky confusion
forever traditional
while often cruel

Fig. 18 Haiku example 1 generated by interactive evolution

The haiku captures a feeling. Confusion is nowadays common, so it can be considered as traditional, and it can also have cruel consequences.

feeling metaphor
beyond a punctuation
lavender blossoms

Fig. 19 Haiku example 2 generated by interactive evolution

This haiku poem can be explained as an attempt to evoke impression of the beauty of cherry blossoms that is captured by another poem (metaphor and punctuation).

equity raises
evil claw releases
blooming daffodil

Fig. 20 Haiku example 3 generated by interactive evolution

This haiku poem is trying to create a picture of ending injustice on the background of spring daffodils.

In some cases, system creates less meaningful poems. Following haiku poem example could be considered as confusing and have very little meaning.

also microwave therapy
kill for ashamed battery
also microwave

Fig. 21 Haiku example 4 generated by interactive evolution

5. COMPARISON

The advantage of interactive evolution approach is the possibility to have personalized poems according the specific user preferences and needs. This approach does not require difficult implementation, since the algorithm does not have domain-specific knowledge. On the other hand, poem generation takes more time.

Approach for poetry generation that uses models is more difficult to implement, because it has to have certain knowledge about poetry. However, the generation of poem is very quick.

5.1. Formal Criteria

During development of applications, it was discovered that not all of the poems in haiku corpus comply with typical haiku syllable pattern. This causes problems in application of interactive evolution. The approach relies heavily on the haiku corpus- that it consists of valid (from formal point of view) haiku poems. When the verse from corpus does not have the right syllable count, it is reflected into the output poem.

Application for haiku generation, that uses sentence models, leverages syllable division algorithm when creating haiku models and thus makes sure the syllable count will be correct.

5.2. Content Criteria

In order to comply with content criteria, haiku corpus consisting of large amount of poems written by human authors is used in both approaches.

Both approaches use the corpus as a source for elementary building blocks (verses in case of evolution, words in case of sentence models) of the poems and both are creating poems by combining the blocks.

The reason for using haiku corpus is to make sure that the basic building blocks will be haiku-specific. This means, they will reflect and contain typical haiku content (which is nature and emotion expression).

5.3. Meaning Criteria

The main reason for creating confusing poems with interactive evolution is mixing poems with different topics together in cross-over. When two poems with different topics and/or with different emotion were selected to cross-over, confusing and sometimes absurd poems were created. These would be eliminated by human evaluation. Since the population has very little individuals and the number of generations also is not high, the system does not have the possibility to converge so quickly in such big search space.

For this reason, sometimes neural network trained on data obtained from human testers are used to be the fitness function evaluating the poems. It also helps with speeding up the evaluation process. For example, system POEVOLVE [15] uses this approach.

The reason, why sometimes less meaningful poems were generated with using poem models is that the approach does not take into account the meaning of words. Words to fill the poem model are selected in regard to their part of speech and syllable count to fulfil the formal haiku criteria. However, the approach does not provide any further knowledge on how to select words into poem, so that it would conform to meaning criteria as well. It is easier to combine words than whole verses. This is the reason why the poem models are more successful in creating poems.

6. CONCLUSION

In overall evaluation, the approach with using poem models has better results. The reason is its focus on words, rather than verses, when creating new poem.

The applications will be later made accessible to wide public in order to evaluate and test the performance of poetry generator with real users and continuously improve its results.

REFERENCES

- [1] KUKICH, K.: Design of a knowledge-based report generator, Proceedings of the 21st annual meeting on Association for Computational Linguistics, Association for Computational Linguistics, 1983.
- [2] GOLDBERG, E. – DRIEDGER, N. – KITTREDGEM, R. I.: Using natural-language processing to produce weather forecasts, IEEE Expert 9.2(1994), pp. 45-53.
- [3] SAUPER, Ch. – BARZILAY, R.: Automatically generating Wikipedia articles: A structure-aware approach, Proceedings of the Joint Conference of the 47th Annual Meeting of the ACL and the 4th International Joint Conference on Natural Language Processing of the AFNLP, Volume 1 - Volume 1. Association for Computational Linguistics, 2009.
- [4] MEEHAN, J. R.: TALE-SPIN, An Interactive Program that Writes Stories, IJCAI, Vol. 77, 1977.
- [5] BRINGSJORD, S. – FERRUCCI, D.: Artificial intelligence and literary creativity: Inside the mind of brutus, a storytelling machine, Psychology Press, 1999.
- [6] HALL, J. D.: Popular Prosody: Spectacle and the Politics of Victorian Versification, NINETEEN CENT LIT, Vol. 62, No. 2, 2007, pp. 222-249.
- [7] Random Poem Generator, [online, cited 20.12.2016]. Available at <<http://poetry.namegeneratorfun.com/>>.
- [8] ThinkZone Poem Generator, [online, cited 20.12.2016], Available at <<http://thinkzone.wlonk.com/PoemGen/PoemGen.htm>>.
- [9] NSA Haiku Generator, [online, cited 20.12.2016], Available at <<http://www.nsahaiku.net/>>.
- [10] Rzepka, R. – Araki, K.: Haiku generator that reads blogs and illustrates them with sounds and images, Proceedings of the 24th International Conference on Artificial Intelligence, AAAI Press, 2015.
- [11] HRONCOVÁ, M. – MACHOVÁ, K.: Generating Haiku Poems Using Methods of Artificial Intelligence, Proceedings of the 10th Workshop on Intelligent and Knowledge oriented Technologies, WIKT 2015, Košice, 2015, pp. 96-102, ISBN 978-80-553-2271-1.
- [12] Aha Poetry, [online, cited 10.03.2016], Available at <http://www.ahapoetry.com/aadoh/h_dictionary.htm>.
- [13] Daily Haiku, [online, cited 10.03.2016]. Available at <<http://www.dailyhaiku.org/>>
- [14] Talk English Basic Vocabulary, [online, cited 10.03.2016], Available at <<http://www.talkenglish.com/vocabulary/top-2000-vocabulary.aspx>>.
- [15] LEVY, R. P.: A computational model of poetic creativity with neural network as measure of adaptive fitness, In: Proceedings of the ICCBR-01 Workshop on Creative Systems, 2001.

Received September 8, 2016, accepted February 13, 2017

BIOGRAPHIES

Miroslava Hrešková was born on 16.08.1992. In 2015 she defended her diploma thesis in the field of Artificial Intelligence; her thesis title was “Design of System Using Interactive Evolutionary Computation in Social Robotics“. Since 2015 she is working as a software engineer and is a distant PhD. student of Intelligent Systems at the department. Her scientific research is focusing on computational creativity and poetry generation.

Kristína Machová was born on 16.02.1961. She graduated (MSc.) in 1985 at the Department of Technical Cybernetics at the Technical University of Košice. She defended her PhD thesis in the field of machine learning in 1996. She works as an associate professor at the Department of Cybernetics and Artificial Intelligence of the Faculty of Electrical Engineering and Informatics at the Technical University of Košice since 2009. Her scientific research focuses on machine learning, automatic classification of text documents, information retrieval, texts processing, etc. In addition to this, she also investigates the questions related to the social web and various methods for the conversational content analysis.