

# Linguistic Model Propositions for Poetry Retrieval in Web Search

Magdalena Granos and Aleksander Zgrzywa  
 Wrocław University of Technology  
 Wybrzeże Wyspińskiego 27, 50-370 Wrocław, Poland  
 E-mail: magdalena.granos@pwr.wroc.pl

**Keywords:** linguistic models, poetry retrieval, information retrieval, web search

**Received:** January 23, 2012

*The paper investigates the linguistic models for poetry retrieval purposes and raises the importance of access to advanced exploratory search engines, which, according to the Maslow's hierarchical pyramid of needs, address the needs of higher order associated with self-fulfilment and satisfaction of informational-cognitive aesthetic ambitions of non-verbal transmission. This task statement is formulated in the context of a new scientific discipline – cognitronics. The work focuses on filling the information needs of sublime poetry as a form of structurally categorized pattern, which is an important indicator to determine the search methods of poetry based on adaptation of linguistic models. The issues addressed refer to probabilistic models that allow for predicting the occurrence of words in a sentence on the basis of distance functions relatively to similarity of words and phrases, as well as by k-nearest neighbours strategy, and frequency of words relatively to rankings. Moreover, the paper aims at showing the optimal performance of linguistic models in search of the most effective methods of poetry retrieval. During these searches, the relevant information retrieval models reveal their role in terms of new opportunities for Internet search engines into the formative process of education and creation. Accordingly, the paper attempts to explore ways of increasing web-search efficiency to make future research flexible, yet precise, in interpretation of queries.*

*Povzetek: Opisana je gradnja in uporaba lingvističnih modelov za iskanje pesmi na spletu.*

## 1 Introduction

Amid escalating resources of complex data in website network, the adaptation of search tools for information from the extensive resources of knowledge, implies the inevitability of working towards increasing power search systems. Globalization process of transformation in socio-economic context, calls for smooth availability and flowing information, as well as an operative technology to provide efficient multi-contextual information services for efficiently informed society. Widely and easily accessible information is a common and inexhaustible wealth of human imagination and intellectual achievements, which are inalienable in consciousness in comparison to the resources of derivatives of life.

The information requirement is the space of human knowledge, experienced consciously as a query, leading to look for relevant reply. As far as a user of information system is considered, a measure of meeting the need for information involves the assessment of user knowledge about information systems and services. It also results from the size of a subset of user's population, who positively evaluate a system. Eventually it concerns the extent to which people use different types of services, resources and information collections [17].

During last decades, especial after the birth of the Internet, people in developed countries have been exposed to permanently increasing informational and time pressure, have been forced to take much more

decisions in the same time as they used to do before. As a result, people have now much less time for the spiritual-emotional development, and it is one of negative, broadly observed shifts in modern information society. That is why a very significant problem for the humanities, computer science, and web science is to find new ways of supporting spiritual-emotional development of the personality in the information society.

The analysis of the mentioned and some other negative shifts caused by the progress of information and communication technologies and globalization processes, the desire to make something constructive for compensating the discovered negative shifts have caused the birth of a new scientific discipline called Cognitronics [8, 9].

It should be underlined in the context of this paper that Cognitronics formulates a new, large-scale goal for the software industry and Web science: “to develop a new generation of culture-oriented computer programs and online courses (in the collaboration with educators, linguists, art historians, psychologists) - the computer programs and online courses intended for supporting and developing positively-oriented creativity, cognitive-emotional sphere, the appreciation of the roots of the national cultures, the awareness of the integrity of the cultural space in the information society, and for supporting and developing symbolic information

processing and linguistic skills, associative and reasoning abilities of children and university students [8].

Along with the growing expectations of information, there is a pressing need for advanced exploration capabilities of Internet search engines, which take into account not only the wide spectrum of phrases, determined by semantics terms from surrounding reality in context of accompanying circumstances. According to the hierarchical pyramid of needs by Maslow as a consequence of once fulfilled expectations, owing from the vital functions, aspirations of a higher order arise, associated with self-fulfilment and satisfaction of informational-cognitive aesthetic ambitions [14]. Poetry, pervading the most subtle areas of human sensitivity, fills the most sophisticated information needs, and becomes an area of particular desirability to find information about non-verbal lore.

## 2 Concept of Information Need

The concept of information need concerns the gap in a person's knowledge. The information request is expressed as a query asked for relevant sources. An inquirer obtains an answer from an information system.

The substance of an information need is determined during an information interview, giving the recommendations of relevant sources.

In case of a user of information system, a measure of meeting the need for information is associated with:

- an assessment of user knowledge about information systems and services;
- the size of a subset of the population of users, positively evaluating a system or its components;
- the extent, what people use different types of services, sources and collections of information.

We live in the Age of Information and rapid communication. Intensifying the size and flow of information, determines more sophisticated expectations.

The more accessible information becomes, the more possessive and irresistible become expectations for ubiquitous and incessant access in a variety of situations [13] in consequence of claim for advanced culturally, fanciful and subtle areas of information with ambiguous contexts.

Following the Maslow's pyramidal hierarchy of needs, directly from the bottom of basic physiological needs, successively across safety, belonging, esteem, up to the peak of the self-actualization, meant by creativity, intuition, identity and purpose, once the basic needs are met, information becomes more valuable in personal growth.

Such tendencies have continuously claimed research and development of the information and search systems, to fulfil the desire for the most advanced and sophisticated culturally spaces of information, such as fanciful and subtle field of poetry based on nonverbal needs of a user.

Mehrabian's communication model, where only 7% are the words of verbal communication, while 38% and 55%, respectively, constitute intonation and expression of non-verbal communication, shows the validity of new

trends development of search mechanisms [16]. They have intuitively interpreted the meaning of messages, which are expressed in poetry through stylistic means, such as onomatopoeic words, accent, rhythm, rhymes, expression of feelings, mood, metaphor, the spatial organization of words and thematic motifs as well [3].

Although being non-verbal in communication, poetry, like musical notation, also appears as a structured form, that is operating according to the categorized models, such as the rhythmic pattern, metric foot, rhyming pattern, versification pattern, alliteration, which are effectively used to develop a search method, submitted in consequence of building a new linguistic model based on structural patterns of poetic works, and by adapting some linguistic models [7].

The process of searching for poems in web browsers, concerning the semantic and syntactic patterns of poetry means, makes the essence of the poetry retrieval.

## 3 A Philosophy of Language Approach

Philosophy of language has generally been concerned with language and meaning on focus of empirical investigation in linguistics as scientific study of natural language. This is generally referred in philosophy to as the 'Linguistic Turn'.

Reviewing the philosophic configurations of the language in the twentieth century in the context of life-meaning aspects of special significance are the Chomskian linguistics, which are particularly relevant in a perspective derived from the Cosmonomic Philosophy of Dooyeweerd. Noam Chomsky perceives language as a supervening property of a certain system found in human brain - he calls this a Language System - which development's determines individual and its own internal language as a principal object of study [5].

Herman Dooyeweerd extends his idea of Multiple Rationalities through a new philosophic approach to such issues as meaning, knowledge, being, time, functioning, normativity, theory, practice or social structures, providing grounds for treating life-meaning and the life-world [6], where all human activity is multi-aspectual. Dooyeweerd believes that the kernel meaning of various aspects can be grasped by our own intuition. For example, a concept of justice is imprecise and often divisive but it can also be understood very intuitively.

Similarly, reflecting on aspects of natural laws governing Cosmos, one can intuitively extend certain imperatives into various concepts, faiths or languages. Thus, reflecting on aspects of poetic language, aside of often intuitive and unquantifiable beauty, we can also assess a poem in normative forms, since a poem functions formatively (it has a structure), lingually (it is writing), aesthetically (harmony of style and play of words), juridically (copyright), and so on. Certain aspects are more important than others: the poem can still be a moving expression of aesthetic writing, even without copyright.

The aesthetic and lingual aspects seem particularly important for a poem, though in different ways. The

lingual aspect refers to the ‘material’ of which the poem is made. In turn, the aesthetic aspect refers to the type of normativity that determines its quality [2].

The ideas of philosophy of language demonstrate how a perspective in a scientific area might be developed to for incorporating life-meaning based methods into contemporary research.

The philosophical support points that language meaning is rooted in participating in human ways of life, and only human beings can function as subjects in the lingual and post-lingual aspects. What this means is that the human user needs to be involved intimately in the processing of text. ‘I’ll build a house right at eleven o’clock’ is only understandable once one knows the life-meaning of communities that employ this idiom.

Despite some limitation, then it is likely that better rule sets can be compiled at least to detect and model the life-meaning issues with aim to improve the accuracy, namely the number of words correctly disambiguated, ipso facto preventing from word sense ambiguity, which is one of the main causes which affect the retrieval performance in the field of information retrieval and natural language processing, on account of the polysemies (when one word may have different meanings under different contexts) and synonymies (when different words may have the same meaning).

#### 4 Nonverbal Aspects of Information Need

The concept of information need concerns the gap in a person's knowledge, urgently followed by craving for satiation of the adequate reference. The information request is expressed as a query asked for relevant sources.

An inquirer obtains an answer from an information system on a path of some levels of exploration. This starts with forming the actual, but still unexpressed need for information to evaluate then a rational statement and unambiguous description of the doubts and finally ask a question as presented to the information system.

To understand properly and respond adequately to users, the semantic extraction has become the major challenge in the development of the semantic web. Unavoidably, the complex process of extracting semantic information from natural language documents generates simultaneously particular difficulties. The semantic extraction from documents such as poetry is even more difficult due to its composite metric patterns, miscellaneous styles, ambiguity and elusiveness as well as symbolic meanings.

Arising from the philosophy of language, in consequence of the innate facility for language possessed by the human intellect, the natural language documents often appeal for nonverbal communication, as a means to reinforce and complement the message. Via its functions of complementing (adding extra information to the verbal message), contradicting (when our nonverbal messages contradict our verbal messages), repeating (used in order to emphasize or clarify the verbal message), regulating (serves to coordinate the verbal dialogue between

people), substituting (occurs when a nonverbal message is transmitted in place of a verbal message) [11], the nonverbal communication is usually understood as the process of communication through sending and receiving wordless messages, however the written texts have also their nonverbal elements, such as spatial arrangement of words or emotional expression of feelings.

There are also areas where the verbal and nonverbal means of communication overlap, like in poetry, which carries rhyming patterns, rhythmic regularity in versification and metaphors with capability to express the inexpressible and assimilate dissimilarities. Thus, for poetry, with its metrical parallelism of lines and the phonic equivalence of rhyming words, metaphor, is the line of least resistance and, consequently, the study of poetical tropes is directed chiefly toward metaphor [10].

Traditionally, metaphor was a term in rhetoric, a term which referred to a purely linguistic figure, defined in the Aristotelian manner as the transfer of a name from one thing to another [4]. But with the advent of Structuralism, metaphor has acquired a greatly enlarged, granted that every realm of human culture can be construed as a type of language, what brings to light properties and relationships, which had hitherto gone unnoticed.

How far the nonverbal aspects of language can grasp and how deeply they can touch the ambiguity of the context, a simple task of text retrieval may visualize how it is used within linguistic context of a local culture in the Tatra region in Poland, a region of great natural beauty and folkloric originality, but completely novel in exploratory experience, including a peculiar local dialect, not readily comprehensible even to most of Poles. A locution ‘I will build a house right at eleven o’clock’ inclines towards a meaning of the time, when it will be raised, whereas the idiom practically concerns the building a house facing South. A user who wished to find texts that refers to the building a house facing South would completely miss those, that included ‘To build a house right at eleven o’clock’, unless the prevalence of retrieval methods respecting the nonverbal aspects of information need.

#### 5 Structural Patterns of Poetic Works

The principle problem of the poetry retrieval is to find out the adequate methodology to define precise, relevant and accurate enough, web search model, able to handle with such obstacles like word sense ambiguity on account of the polysemies (when one word may have different meanings under different contexts) and synonymies (when different word may have the same meaning).

According to Chomsky's, the ability to learn and use language is an innate property, that generates the words in a hierarchical format of branched tree, called the structure of the syntax [5]. Chomsky's revolution shows that the internal capacity of learning, processing and building syntax is universal to the species, since we are

born with sensitivity to this type of structural system, functioning on the principle of tree [15].

The nonverbal elements in real life can be, for instance, voice intonation, feelings expression by laughing, crying, sighing, as well as the elements of the physical

environment, like lighting, sounds, smells, whereas the nonverbal elements in poetry are expressed through stylistic means, such as onomatopoeic words, accent, rhythm, rhymes, expression of feelings, mood, metaphor, the spatial organization of words and thematic motifs as well.

As an example of versification pattern, appears alliteration, what is a repetition of the initial consonantal sounds of syllables and words, repeating the same letters and syllables at the beginning of words in the verse, the next verse or sentence. We can say that alliteration is like rhyming, but with alliteration the rhyming comes at the front of the words, instead of at the end.

As an example of rhythmic pattern, appears metric foot, what is a unit of measurement in poetry, while meter refers to the repeating pattern of stressed and unstressed syllables in the lines of a poem, just like in a screened example, where each unstressed syllable occurs in italics and each stressed syllable in bold.

The structural patterns of poetry are determined by, inter alia:

- literary genres: lyrics (ballad, epigram), epics (ode, elegy, pastoral, song, sonnet), drama (tragedy, comedy, monologue);
- literary eras: Middle Ages, romanticism, contemporary literature;
- mood (joy, sadness, happiness);
- thematic motifs (love, farewell, nature).

Application of methods for poetry retrieval, based on structural patterns of poetic works and adaptable linguistic models, will allow the efficient obtaining of relevant answers to questions, concerning the poems. Effective developing of the methods of poetry retrieval in web search can flow owing to adaptation of selected linguistic models on the structural patterns of poetic works.

## 6 Linguistic Model of N-gram Text Categorization

The Linguistic Model of N-gram Text Categorization is a probabilistic model that allows predicting the next elements in a sentence.

N-gram is a subset of sentences, composed of  $n$ -elements of the sentence, which may be letters, characters, syllables, words or pairs of elements mostly of sizes from 1 to 5. The most elementary uni-gram model rejects the conditional context and defines each term separately, hence usually used to find information in case of structural complexity documents, whereas the second size bi-gram model determines the occurrence of the words preceding [12].

The probability  $P(W_n|W_{n-1})$  of  $W_n$  (where  $W$  is a word and  $n$  is a position in a sentence of  $W_n$  word occurrence), conditioned by the occurrence of a  $W_{n-1}$  preceding word, is equal to the ratio of the probability of co-occurrence of two words  $P(W_{n-1}, W_n)$  to the probability of  $P(W_{n-1})$  preceding words (equation 1).

$$P(W_n | W_{n-1}) = \frac{P(W_{n-1}, W_n)}{P(W_{n-1})} \quad (1)$$

The functions of distance between the document and category profiles are determined, according to the model, based on quantitative Zipf law, where word frequency is inversely proportional to the ranking, what means that the higher the rank, the lower frequency.

In data flow of the N-gram model of text categorization the following steps take place (Figure 1):

- text is divided into separate tags, consisting only of letters, an apostrophe and a space before and after the characters. Numbers and punctuation marks are discarded;
- each tag is scanned, generating all n-grams in size from 1 to 5;
- using the mechanism of collision each n-gram is given its own counter;
- all the n-grams are counted;
- N-grams are sorted in reverse order to the number of occurrences;
- the resulting file is a profile of n-gram frequencies for a document;
- the profile distances are measured;
- for each n-gram in a document profile, we find the equivalent in the profile category and count how far away. (For example, the n-gram "ING" is a ranking of 2 items in the document, but at 5 in the category, so it has a ranking of 3, after measuring the distance. If n-gram as "ED" does not appear in the profile category, with the value of the maximum distance.);
- the sum of all distance values for n-grams is a measure of the distance from the category of the document;
- to specify the minimum distance, the measure of distance for all profiles of categories to profiles of documents is taken and the smallest of them is a result.

The benefits of such approach is the ability to work both with short and long documents, also the minimum occupancy requirements of memory and computing, as well as a perfect fit to the texts of noisy sources.

The ability of predicting the next elements in a sentence through the process of measuring a distance between n-grams of the documents can also be effectively adapted to a language model for poetry retrieval, where the sum of all distance values for n-grams is a measure of the distance from the assumed words of the searched poems and the words of the retrieved poems, whereas the smallest of distances give an optimal result of poetry retrieval.

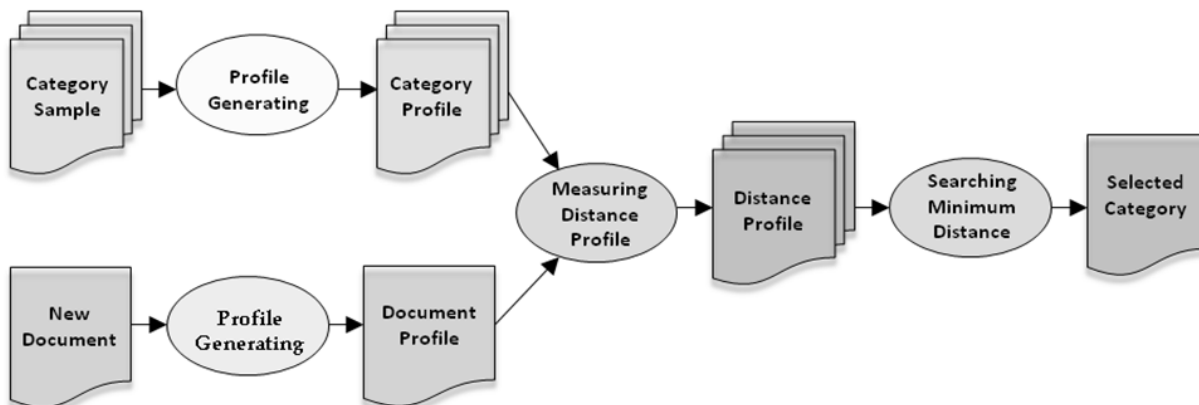


Figure 1: Data flow in the N-gram model of text categorization.

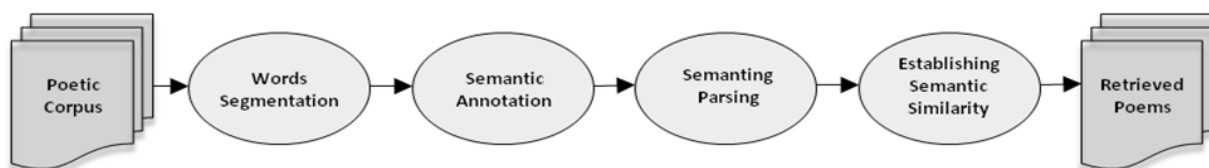


Figure 2: The flow of data in the semantic model of structural similarities.

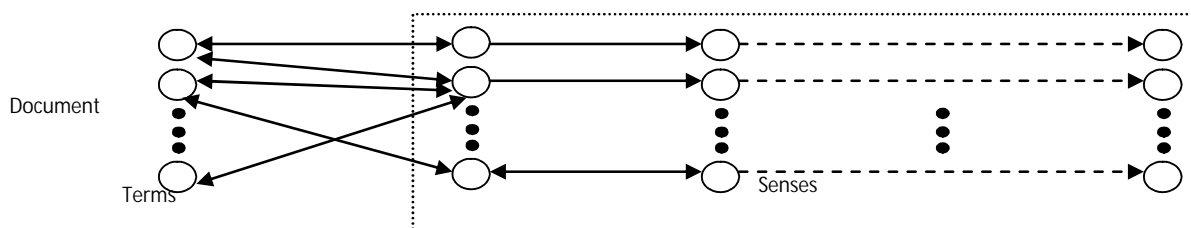


Figure 3: Language Sense Model for Information Retrieval.

### 7 Linguistic Semantic Model of Structural Similarities

Searching of poetry from the sources of poetic corpuses is done by measuring the scale of the semantic similarities on structural paths of queries and lines of poems on the basis of semantic annotation and parsing (Figure 2).

The Linguistic Semantic Model of Structural Similarities is based on the tree structure, adopted to search for poetry by using the algorithm of structural similarities [18], where:

- the T1 tree consists of n1 nodes;
- C(n1,n2) is the number of subtrees with roots at the nodes n1, n2 (When a semantic category for n1 is different than for n2, then C(n1,n2)=0, where n2, n1 are the final nodes and have the same category, meanwhile C(n1,n2)=sim(n1, n2).);
- K(T1,T2) is a measure of similarity between trees T1 and T2, which is the sum of similarity measure C(n1,n2) in each pair of nodes of trees T1 and T2 (equation 2);

$$K(T_1, T_2) = \sum_{\substack{n_1 \in N_1 \\ n_2 \in N_2}} C(n_1, n_2)$$

$$C(n_1, n_2) = sim(n_1, n_2) \tag{2}$$

$$sim(n_1, n_2) = \frac{1}{4^N}$$

- sim(n1,n2) – is the result of semantic similarity and capital N is a number of branches between n1 and n2 (In case, sim(n1,n2) is equal 1/16, then N equals 2, what means, that there are 2 branches between n1, n2; if sim(n1,n2) is equal = 1/256 then N equals 4, in other words, there are 4 branches between n1 and n2. Eventually, if sim(n1,n2) is equal = 1/4096 then N equals 6, and that means, there are 6 branches between n1,n2 and semantic similarity tends to infinity).

The flow of data in the semantic model of structural similarities shows that poetry retrieval in the sources of poetic corpuses is done by measuring the scale of the semantic similarities on structural paths of queries and lines of poems on the basis of semantic annotation and parsing.

The first step in the process is segmentation of words, when poetic lines of the body are segmented according to the dictionary of synonyms.

The next stage is semantic annotation, when expert annotates every word with the correct semantic code, in 4-leveled taxonomic hierarchy, in which the first level is the most abstract, and the fourth most detailed. Each term in a dictionary of synonyms has a semantic code that

represents a hierarchical classification. For instance, for the word 'general' there is a semantic code AE1004, which is represented by four code elements AE 10 04, where A is the first-level category code for a person, AE means the profession of a persons, AE10 goes deeper for 'military rank' and finally AE1004 is a "specific name of military rank".

Subsequently, each line of poems is being parsed on the basis of the metric pattern during the semantic parsing.

Eventually, the semantic similarity is being established through measuring the similarity of the semantic structure of each line of poems to the semantic structure of queries. Consequently, finding poetry is done by using the algorithm of structural similarities. The lower distance indicates the higher similarities.

## 8 Language Sense Models for Information Retrieval

Due to the vision of semantic web of universal flexibility to fulfil the requests of users, the extraction of the semantic information from natural language documents has become one of the major challenges to undertake. A great deal of work has been done on drawing word senses into retrieval to deal with the word sense ambiguity problem, but with few positive results. And yet the first significant accomplishments of such scientific achievement give the language sense model (LSM).

A document model generating queries can be used either to recognize or to generate strings. The full set of strings that can be generated is called the language of the automaton, consisting of nodes, capable to exchange information over a communications channel. If each node has a probability distribution over generating different terms, then we have a language model [12], which is a function that puts a probability measure over strings drawn from some vocabulary.

In the LSM the model generates the probability of a given query from both document's term and sense representation (Figure 3).

The LSM combines the terms and senses of a document seamlessly through an expectation-maximization (EM) algorithm for data augmentation and optimizing a likelihood function, what lets overcome such problems of research work related to language model as data sparseness (the existing smoothing methods can be applied easily on both terms and senses to solve the data sparseness problem) and term dependency (the query independent assumption can be relaxed to a certain extent as the terms and senses in the LSM depend strongly on each other) [1] (equation 4).

$$P(q|d) = \prod_{i=1}^m ((1-\lambda)P(q_{t_i}|d_{t_i}) + \lambda P(q_{s_i}|d_{s_i})) \quad (4)$$

Retrieval on lexical collections shows that the LSM outperforms the traditional language model for both medium and long queries, however, not significantly on short queries, as there are less nouns and verbs to be

disambiguated for short queries as well as it's much harder to disambiguate the short queries because of the sparse context.

The forthcoming attempts anticipate the hierarchical smoothing using more relations of lingual database groups words into sets of synonyms with semantic relations between them.

The purpose is to produce a combination of dictionary and thesaurus that is more intuitively usable, and to support automatic text analysis and artificial intelligence applications throughout the evaluation of the LSM on more corpora.

## 9 Conclusions

Usage of poetry retrieval methods, based on structural patterns of poetic works and adaptable linguistic models, including the distance measure of words and phrases similarities about non-verbal meaning as well as building new linguistic models and using structural patterns of poems, will allow the efficient acquisition of response to questions, which concern poetry retrieval.

Exploring the websites databases of poetry works, based on linguistic models, shows that the present systems of poetry retrieval reveal weaknesses within identification of the contextual ambiguity and under accuracy of information retrieval.

Validity of information retrieval in context of non-verbal requires of users emphasizes this year's report by the

International Telecommunications Union, which shows that yet 10 years in the whole world was less than 400 million Internet users, while today there are exactly 2.08 billion, which means that about 1/3 of humanity benefits from Internet resources, including the 950 million users of mobile broadband, and 555 million from fixed broadband with a steady upward trend of potential target groups and relative areas of application, carrying the indisputable benefits of thorough exploration to personal growth and flowering of modern education and information society.

It can be concluded that poetry retrieval might be both possible and beneficial, giving opportunity of effective efforts towards thoroughgoing web collection research.

## 10 References

- [1] Bao S., Zhang L., Chen E., Long M., Li R., Yu Y. (2006). LSM: Language Sense Model for Information Retrieval, Lecture Notes in Computer Science, pp. 97-108.
- [2] Basden A., Klein H. K. (2008). New research directions for data and knowledge engineering: A philosophy of language approach Introduction to Information Retrieval, ELSEVIER/Data & Knowledge Engineering, Vol. 67, No. 2, pp. 260–285.
- [3] Bojar B. (2002); Słownik encyklopedyczny terminologii z zakresu języków i systemów informacyjno-wyszukiwawczych.

- [4] Bredin H. (1990), *Metaphor, Literalism, and the Non-Verbal Arts*, Springer Netherlands, Vol. 20, No 3, pp. 243-341.
- [5] Chomsky N. (2000). *New Horizons in the Study of Language and Mind*. Cambridge University Press, pp. 1-195.
- [6] Dooyeweerd H. (1984). *A new critique of theoretical thought*. Paideia Press, Vol. 1-4, pp. 1-2205.
- [7] Donald R. B. (2000). *Nonverbal Poetry: Family Life-Space Diagrams*; *Journal of Poetry Therapy*, Vol. 14, No. 3, pp. 159-167.
- [8] V. Fomichov and O. Fomichova (2006). *Cognitonics as a New Science and Its Significance for Informatics and Information Society. Special Issue on Developing Creativity and Broad Mental Outlook in the Information Society (Guest Editor Vladimir Fomichov)*, *Informatica. (Slovenia)*, 30 (4), pp. 387-398.
- [9] O. Fomichova and V. Fomichov (2009). *Cognitonics as an Answer to the Challenge of Time. Proceedings of the International Multiconference Information Society- IS 2009, Slovenia, Ljubljana, 12 – 16 October 2009. The Conference Kognitonika/Cognitonics. Jozef Stefan Institute, 2009, pp. 431-434; available online at <http://is.ijs.si/is/is2009/zborniki.asp?lang=eng>.*
- [10] Jakobson R., Halle M. (1956). *Fundamentals of Language*. The Hague & Paris: Mouton, pp. 55-82.
- [11] Malandro, L. A, Barker L. L., Barker D. A (1989). *Nonverbal Communication*. Reading MA: Addison-Wesley.
- [12] Manning C. D., Raghavan P., Schütze H. (2008). *An Introduction to Information Retrieval*; Cambridge University Press, pp. 237 –252.
- [13] Marchionini G., White R. W. (2009), *Information-Seeking Support Systems*, *IEEE/Computer*, Vol. 42, No. 3 (pp. 30-32).
- [14] Maslow A.H. (1943). *A Theory of Human Motivation*; *Psychological Review*, Vol. 50, No. 4.
- [15] McWhorter J. (2008). *Understanding Linguistic: The Science of Language*; The Teaching Company.
- [16] Mehrabian, A. (2002). *Silent messages. Nonverbal communication*; Reedition Publishing, Springer-Verlag, New York.
- [17] Reitz J. (2004). *Dictionary for Library and Information Science*; *Libraries Unlimited*, pp. 357.
- [18] Shu-Lei Chen W. (2004). *Semantic Structure Extraction and Retrieval of Chinese Poetry*, National Tsing Hua University, pp. 1-49.

