Conceptual Semantic Relationships for Terms of Precalculus Study

VELISLAVA STOYKOVA
Institute for Bulgarian Language - IBL
Bulgarian Academy of Sciences - BAS
52, Shipchensky proh. str. bl. 17, 1113 Sofia
BULGARIA
vili1@bas.bg

MAYA MITKOVA
Gulf University for Science and Technology
P.O. Box 7207
Hawally 32093
KUWAIT
mitkovam@gust.edu.kw

Abstract: In this work, we present the results of research using the techniques for semantically-oriented statistical search for extracting mathematical terms for precalculus with applications to education, terminology and ontology. We offer the combination of statistically-based techniques incorporated in the software for extracting keywords, collocations and co-occurrences as the useful tools for teaching courses development, domain terminlogy definitions, and building ontologies.

Key–Words: Basic Science in Engineering Education, New Technologies in Education, Research and Education, Computers, Internt, Multimedia in Engineering Education, Research and development in Engineering Education.

1 Introduction

The engineering education includes compulsory courses of mathematics at a different extend. However, the university content and proportion of different mathematical subjects in courses for university mathematics may vary depending on the national educational standards and mostly depends on the secondary and university level education. But traditionally, it includes teaching and mastering general topics of precalculus.

Thus, mastering basic mathematical concepts is a key teaching activity for the improvement of students' mathematical ability and development of engineering creativity. The problem of designing courses including teaching precalculus is often considered as a problem of the way the different basic mathematical concepts and their order are transposed and introduced to students in the course outline.

The problem of creating practical study course materials can be, also, addressed as a process of creating a mathematical terminological conceptual knowledge hierarchy which defines the organization and the structure of teaching precalculus course material.

Further, we are going to present and discuss the basic principles and the results of a corpus-based research and analysis of three different web-based open-source course materials for precalculus study based on the techniques of computer-supported extraction of lexical semantic and terminological database design.

2 Artificial intelligence approaches to terminology

Artificial intelligence (AI) methodogies and techniques, recently developed, have influenced traditional terminlogical practice in many ways, resulting in a wide range of real applications. Thus, the use of electronic text corpora of different genres instead text archives used in the past, speeded significantly the production of different terminlogical reference sources like thesauri, specialized dictionaries or encyclopedias offering fast production of updated editions by using proper software tools assisting various types of search.

However, the terminolgy still is considered as the area with its own scope, which is different than that of lexicography, and is mostly focused on describing, creating and structuring the domain conceptual knowledge by defining the domain conceptual semantic relationships. It, also, has wide range applications with successful results in education by offering approaches to structuring the teaching material or preparing the related school books.

At the same time, terminological reference sources are often used as a reliable base for creating the electronic dynamic terminological conceptual knowledge hierarchical on-line database.

Thus, we are going to present and analyse the results of a corpus-based research with applications to education with conceptual semantic relations definitions in the domain of precalculus by using semantically-oriented statistical search to extract the domain conceptual semantic relations.

2.1 Methodologies for term acquisition

AI methodological approaches to terminology use the general term *domain knowledge* to refer to terminological analysis. However, the language of a specific domain, is in fact, the language with the same grammar but well-defined domain conceptual semantic relations, which open the question for multilingual interpretations.

Thus, the AI approaches to terminology mostly include the computer-assisted methods and techniques for terms extraction and for defining their internal semantic relations.

2.2 Computer-assisted terminology acquisition

The computer-assisted terminology extraction has been the most successful technique recently developed and applied for the creation of higly structured and semantically-oriented lexical reference sources like dictionaries, thesauri, etc. It is based on the extensive use of electronic text corpora assisting various types of search procedures.

There are two general types of corpus-based applications - using rule-based search techniques (mostly encoding grammatical relations like inflection [14, 16], syntax, etc.) and using statistically-based search techniques (mostly for extracting the lexical semantic relationships [13, 10] to investigate the word behaviour in large-scale electronic text collections. Some hybrid systems allow the combined application of both techniques, depending on the specific research tasks.

Further, we are going to present the combination of different statistical search techniques to define the semantics of some basic mathematical concepts for precalculus by using the statistically-oriented search.

3 The basic concepts in precalculus for educational use

The mathematics, itself, can be regarded as a special kind of symbolic language but, nevertheless, it uses natural language both for definitions and for explanations [15]. Mathematical texts, and related teaching courses, include numbers, formulas, tables, pictures, and graphics by means of which the meaning of their semantic content can be fully unterpreted.

In our research, we have used the texts as they appear in the teaching courses but we have analysed and interpreted only content words since they stand for the concepts.

3.1 Mathematical concepts and their relations

Acquiring mathematical knowledge is a complicated process of mastering the basic concepts by revealing their semantic relations in the appropriate order which is aimed to develop the ability for thinking and creativity.

It success depends on various requirements like the appropriate curricula design and a related syllabus, the use of highly successful teaching methods - traditionally used and recently developed (including e-Learning and Learning-oriented methods [1]), the preliminary students' ground knowledge, the common knowledge from related subjects [9], the students gender [8], etc. But in general, it is mostly influenced by the definition and mastering of the basic concepts over which the new knowledge is structured.

3.2 Defining the basic concepts by using word frequency

The general approach we are using for defining the basic domain concepts is similar to that used for the EU Life Long Learning project KELLY (KEywords for Language Learning for Young and adults) where the basic concepts are defined by searching the contrasting electronic multilingual text corpora database including the British National Corpus (BNC) for extracting keywords frequency lists [6].

Word frequency lists are interpreted by psychologists as a basic concepts for knowledge acquisition and understanding. Educationalists hold the view that word frequency lists are playing an important role for learning to read and mastering similarity. In general, word frequency lists usually defines the basic domain knowledge concepts.

4 Examining contrasting corpora for keywords by statistical search

Word frequency lists are widely used for various educational tasks like curriculum development, writing syllabus and preparing tests for quality evaluation. So that, in our research [17] extracting keywords frequency list is a task of prime importance.

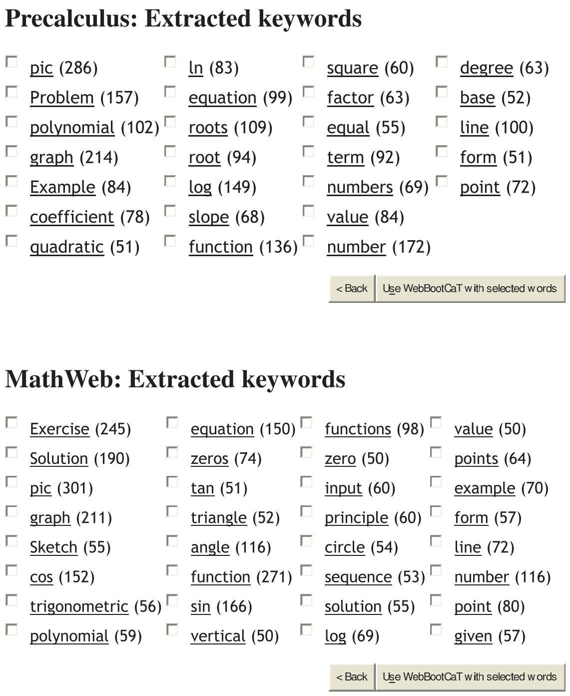

Figure 1: Keywords for MathPre and MathWeb.

Figure 2: Keywords for MathWiki corpus.

4.1 Text corpora and keywords definitions

For our research, we have created three electronic text corpora - MathPre (consisting of precalculus e-course materials given at [18]), MathWeb (consisting of precalculus e-course materials given at [12]) and MathWiki (consisting of precalculus web electronic encyclopedic description which reflects the mathematical presentation of precalculus text materials from Wikipedia [11] and which extensively uses the basic mathematical concepts' relations interpretation introduced in [3]) - of almost 200 000 words. Also, the BNC is used as a standard to compare and interpret the results. In the entire work, we extend the results and the analyses based on and already presented in [17].

There are various statistical approaches to define keywords. In general, most of them define the task for keywords extraction as the retrieval and clustering of statistically similar words [7] and they differ with respect to the statistics used, i.e. with respect to the way they define the semantic similarity.

For the purpose of our analyses, however, we use the statistics incorporated in the Sketch Engine (SE) [4, 5] software for processing corpora which allows the use of elaborated combined semantically-oriented statistical approaches and comparison of the results between several corpora.

The Sketch Engine, also, allows the evaluation of semantic similariry between words based on their grammatical simmilarity with respect to the related inflected word forms or part-of-speech categorization frames which we did not use for our research.

After using the SE standard statistical options for processing our three corpora for keywords definition, we have obtained the results given at Fig. 1 (for MathPre and MathWeb) and Fig. 2 (for MathWiki).

The results represents the basic precalculus concepts like *function(s)*, *numbers*, *polynomials*, *graphs*, *equations*, etc. They differ with respect to the extend and structure of e-course materials used and with respect to the text type of MathPre and MathWeb texts. They, also, reflects the encyclopedic knowlegde presentation structure of MathWiki corpus, which needs a further elaboration to be used for teaching courses.

In general, our keywords frequency lists give the basic mathematical concepts over which the precalculus teaching is structured, however, the proportion, their internal conceptual relations, and their order have to be clarified by using further statistical research to define their semantic relations. Further, we are going to define the semantic relationships for the basic concept *function(s)*.

Precalculus

Page 1 of 7 Go

Next | Last

3, 4 4, and so on. What is a function	? WHEN ONE THING DEPENDS on
- we say that the first is a " function	" of the other. The area of a circle is
other. The area of a circle is a function	of -- it depends on -- the radius.
typically x and y, is called a function	if there is a rule that assigns to each
of y. We then say that y is a function	of x. Thus a "function" must be single
y is a function of x. Thus a " function	" must be single-valued ("one and only
are called the domain of the function	. We say that those are the values for
are the values for which the function	is defined. In the function y = 2x + 3,
function is defined. In the function	y = 2x + 3, the domain may include

Figure 3: Concordances of the word *function* from MathPre corpus.

Corpus: MathWeb

Page 1 of 14 Go

Next | Last

of G. We say that A is a function	of r, but B is not a function
function of r, but B is not a function	of G. In general, in order for
order for a quantity y to be a function	of the quantity x, it must be
decide whether or not G is a function	of B. Exercise 1.1.2 Make up
up a formula in which W is a function	of p Exercise 1.1.3 Make up
Scientific calculators contain function	keys. For example, there is
example, there is a squaring function	on most scientific calculators.
the button for the squaring function	(usually marked with the symbol
when pressing the squaring function	button. If we sometimes got one

Figure 4: Concordances of the word *function* from MathWeb corpus.

5 Extracting semantic similarity relations using statistical measurement

There are different approaches to define the semantic context. In fact, the context can be defined both in logical and in linguistic terms, however, it is highly dependent on particular logical or linguistic theory. At the same time, the so-called context-free grammars have been evaluated as unappropriate tools for natural language processing.

Moreover, in general, grammar analyses are highly language-dependent, instead statistically-based methods which might be used for multilingual applications. Currently, the statistical corpus approaches based on the measurement of word similarity and defining words concordances have been widely used in terminology and lexicography for word sense definition and for definitions of conceptual semantic relations.

There are various statistically-based approached over which a semantically-oriented search procedures can be performed. The related corpus query systems allow great flexibility of statistically-based search for co-occurrences and collocations using different statistical techniques where the context is defined in statistical quantitative terms.

Corpus: MathWiki

Page 1 of 33 Go

Next | Last

initiated by Descartes. A function	, in mathematics, associates one
the argument of the function	, also known as the input, with
quantity, the value of the function	, also known as the output.
known as the output. A function	assigns exactly one output to
given set. An example of a function	is f(x) = 2x, a function which
a function is f(x) = 2x, a function	which associates with every
f(5) = 10. The input to a function	need not be a number, it can
object. For example, a function	might associate the letter A with
describe or represent a function	, such as a formula or algorithm

Figure 5: Concordances of the word *function* from MathWiki corpus.

5.1 Conceptual semantic relations

In general, conceptual semantic relations are regarded as to be of two types - horizontal and vertical. The horizontal (linear) semantic relationships are those of synonymy, anthonymy, meronymy, i. e. showing semantic similarity [13], semantic distance, inclusion (part-of-whole), etc.

The vertical semantic relationships express the relationships of ordering or hierarchy. The vertical semantic relationships are those realised by hyperonymy and hyponymy. All types of semantic relationships can be defined by experting the related contexts through the generation of related word concordances based on the use of different statistical corpus-based approaches [10].

The concordances give all occurrences of the target word in its related context wich is generated by using statistical search [13]. The example concordances for the basic concept *function* received from MathPre, MathWeb and MathWiki corpora are given at Fig. 3, Fig. 4 and Fig. 5, respectively.

Concordances define the context in quantitative terms and a further work is needed to be done to define the semantic relationships by searching for co-occurrences and collocations of the related keyword.

6 Extracting conceptual relations by using co-occurrences and collocations

Concordances and collocations are defined by statistical measurement of the words which are most probably to be found with the related keyword. They assign the semantic relations between the keyword and its

Collocation candidates MathPre

Page 1 Go
Next >

		Freq	T-score	MI	MI3
p/n	polynomial	9	2.904	4.971	11.310
p/n	square	6	2.381	5.151	10.321
p/n	exponential	5	2.226	7.795	12.439

Figure 6: Collocations of the keyword *function* for MathPre corpus.

Collocation candidates MathWeb

Page 1 Go
Next >

		Freq	T-score	MI	MI3
p/n	polynomial	31	5.493	6.215	16.124
p/n	exponential	13	3.562	6.385	13.786
p/n	quadratic	12	3.433	6.822	13.992
p/n	sine	10	3.088	5.421	12.065
p/n	sinusoidal	7	2.624	6.951	12.566
p/n	tangent	7	2.584	5.428	11.042
p/n	linear	6	2.392	5.407	10.577
p/n	rational	6	2.354	4.684	9.854
p/n	cosine	5	2.163	4.942	9.586

Figure 7: Collocations of the keyword *function* for MathWeb corpus.

particular collocated word in statistical terms of probability and (or) frequency. The semantic relations, in general, might be of similarity or of a distance.

The statistical approaches we are using to search for co-occurrent and collocated words are based on defining the probability of their co-occurrence and collocation. We have used the techniques of $T-score$, $MI-score$ [2] and $MI^3-score$ [10] incorporated in the Sketch Engine for processing and searching our three corpora.

Basically for all, the following terms are used: N - corpus size, f_A - number of occurrences of the keyword in the whole corpus (the size of the concordance), f_B - number of occurrences of the collocate in the whole corpus, f_{AB} - number of occurrences of the collocate in the concordance (number of co-occurrences).

The related formulas for defining $T-score$, $MI-score$ and $MI^3-score$ are given at Fig. 9.

Collocation candidates MathWiki

Page 1 Go

Next >		Freq	T-score	MI	MI3
p/n	exponential	65	8.001	7.040	19.085
p/n	inverse	29	5.309	6.152	15.868
p/n	rational	24	4.804	5.689	14.859
p/n	propositional	20	4.437	7.011	15.655
p/n	polynomial	23	4.544	4.253	13.301
p/n	complex	19	3.978	3.517	12.013
p/n	logarithm	12	3.268	4.146	11.316
p/n	trigonometric	11	3.072	3.764	10.683
p/n	increasing	8	2.794	6.357	12.357
p/n	tangent	8	2.721	4.716	10.716
p/n	relation	8	2.719	4.689	10.689
p/n	continuous	7	2.605	6.004	11.618

Figure 8: Collocations of the keyword *function* for MathWiki corpus.

However, the three statistical criteria give different conceptual semantic relationships rang lists but all of them evaluate sucessfully the basic domain conceptual relations.

The most likely collocations candidates words (which are the most frequent collocates) for the keyword *function* are given at Fig. 6 (for MathPre corpus), Fig. 7 (for MathWeb corpus), and Fig. 8 (for MathWiki corpus). The rang list is estimated according to $T-score$ criterion but the results for $MI-score$ and $MI^3-score$ are listed asd well.

Thus, the most frequent collocates semantically related to the concept *function* are *polynomial*, *exponential*, *rational*, *propositional*, *complex*, *logarithmic*,

$$\text{MI-Score} \quad \log_2 \frac{f_{AB} N}{f_A f_B}$$

$$\text{T-Score} \quad \frac{f_{AB} - \frac{f_A f_B}{N}}{\sqrt{f_{AB}}}$$

$$\text{MI}^3\text{-Score} \quad \log_2 \frac{f_{AB}^3 N}{f_A f_B}$$

Figure 9: The statistics used for $T-score$, $MI-score$ and $MI^3-score$ measurement.

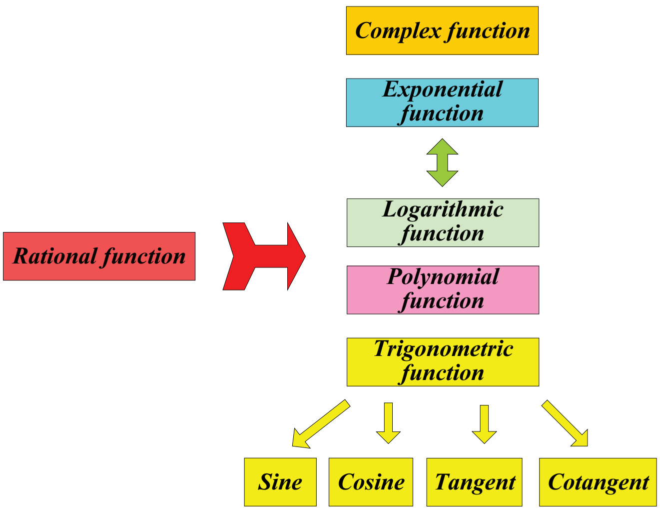

Figure 10: Conceptual semantic hierachy of the keyword *function*.

trigonometric, etc. They express the hierarchical conceptual semantic relationships of the keyword.

Alternatively, the relatively not too frequent collocations like *periodic*, *continuous*, *inverse*, *increasing*, *decreasing*, *real-valued*, *multi-valued*, *positive*, *negative*, etc. represent the attributive semantic relationships of the keyword *function*.

6.1 Conceptual semantic hierarchy

Even the most frequent collocations represent the relationship of similarity, they do not necessarily express the semantic relationship of synonymy. Mostly, they can be interpreted in their hierarchical relationships.

Thus, for our research, we are using such interpretation, and we analyse the *polynomial function* and *exponential function* as the most important concepts to be mastered in teaching precalculus and the *rational function* as the basic concept to start with. The *logarithmic function* can be presented as the inverse to the *exponential function* and the *trigonometric function* can be presented as divided into its subsequent parts *sine*, *cosine*, *tangent*, and *cotangent functions*.

7 Acquisition of domain terminological relations

The above described techniques are fully applicable for domain terminology acquisition. The results achieved, also, can be interpreted by describing and defining the mathematical terms and their internal semantic relations in the domain of precalculus.

Word list

Corpus: MathPre

Reference corpus: preloaded/bnc

Page 1 Go

	user/vstoykova/MathPre		preloaded/bnc		
word	Freq	Freq/mill	Freq	Freq/mill	Score
polynomial	102.0	2065.1	90.0	0.8	1146.4
binomial	30.0	607.4	44.0	0.4	437.0
logarithm	31.0	627.6	86.0	0.8	355.8
parabola	24.0	485.9	42.0	0.4	354.3
y-intercept	16.0	323.9	2.0	0.0	319.2
asymptote	13.0	263.2	8.0	0.1	246.6
logarithms	17.0	344.2	45.0	0.4	246.4
x-intercept	12.0	243.0	0	0.0	244.0
inverses	10.0	202.5	3.0	0.0	198.2
binomials	7.0	141.7	1.0	0.0	141.5
exponential	8.0	162.0	102.0	0.9	85.4
trinomial	4.0	81.0	0	0.0	82.0
calculus	6.0	121.5	108.0	1.0	62.4
logarithmic	4.0	81.0	60.0	0.5	53.4
intercepts	3.0	60.7	30.0	0.3	48.7

Figure 11: Contrasting word frequency lists for MathPre and BNC corpora.

However, the most frequent collocations which express similarity do not represent always the semantic relationship of synonymy. Thus, the *polynomial function*, *exponential function*, and *rational function* are analysed as the most important hyponymic concepts of the very general hyperonym *complex function*. The hierarchical semantic relations of the concept *function(s)* are presented in Fig. 10.

7.1 Defining the domain terms by searching contrasting corpora

As the additional co-occurrences and collocations search techniques, the technique for searching contrasting text corpora for word frequency and comparing the results, also, is considered as a useful tool for defining domain conceptual terms. The technique is used for extracting domain-specific terminology and we apply it for extracting the mathematical terms in the domain of precalculus. Hovever, the technique might be used for improvement of the results of collocation and co-occurrences search techniques.

For our research, we use the three corpora - MathPre, MathWeb and MathWiki - and the BNC as con-

Word list

Corpus: MathWeb

Reference corpus: preloaded/bnc

Page 1 Go

	user/vstoykova/MathWeb		preloaded/bnc		
word	Freq	Freq/mill	Freq	Freq/mill	Score
trigonometric	56.0	1461.1	7.0	0.1	1376.3
polynomial	59.0	1539.4	90.0	0.8	854.7
vertex	35.0	913.2	39.0	0.3	678.4
center	35.0	913.2	83.0	0.7	525.5
parabola	24.0	626.2	42.0	0.4	456.4
cosine	23.0	600.1	36.0	0.3	455.1
directrix	17.0	443.6	1.0	0.0	440.6
ellipse	25.0	652.3	76.0	0.7	389.5
hyperbola	17.0	443.6	21.0	0.2	374.5
asymptote	14.0	365.3	8.0	0.1	341.9
y-axis	13.0	339.2	5.0	0.0	325.7
tangent	23.0	600.1	100.0	0.9	317.8
exponential	22.0	574.0	102.0	0.9	301.2
asymptotes	12.0	313.1	9.0	0.1	290.8
arcsin	11.0	287.0	0	0.0	288.0
intersects	11.0	287.0	25.0	0.2	235.5
quadratic	15.0	391.4	78.0	0.7	231.5
endpoints	9.0	234.8	9.0	0.1	218.3
cotangent	8.0	208.7	0	0.0	209.7
x-intercept	7.0	182.6	0	0.0	183.6
arctan	6.0	156.6	0	0.0	157.6
logarithmic	8.0	208.7	60.0	0.5	136.6
polynomials	8.0	208.7	62.0	0.6	135.1
logarithm	9.0	234.8	86.0	0.8	133.5
arccos	5.0	130.5	0	0.0	131.5
secant	5.0	130.5	2.0	0.0	129.2
intercepts	6.0	156.6	30.0	0.3	124.3

Figure 12: Contrasting word frequency lists for MathWeb and BNC corpora.

trasting corpus. We contrast the search results of the most frequent collocated words for the concept *function* in every one of our three specialized corpora to BNC (which is accepted as a standard corpus). The MathPre, MathWeb and MathWiki are considered as specialized corpora since they are thematically oriented to be in the domain of precalculus.

Nevertheless, there are some differences between them. In general, the MathPre and MathWeb consist of e-course materials already prepared to be used for teaching purposes. Whereas, the MathWiki corpus is a specialized web encyclopedic knowledge description which is prepared to be used mostly as reference source.

The first page results of contrastive word frequency search of MathPre and BNC are given at the Fig. 11. The words which have very low or zero frequency in the BNC are more likely to be the candidates terms. Consequently, the words which have relatively high frequency in the MathPre corpus are also more likely to be the candidates terms. Thus, the words like *polynomial*. *trigonometric*. *asymptote*, *binomial(s)*, *irrational(s)*, etc. are more likely to be terms candidates.

Alternatively, the results of contrastive word frequency search of MathWeb and BNC are given at the Fig. 12. The words like *trigonometric*, *polynomial*, *parabola*, *directrix*, *ellipse*, *hyperbola*, *asymptote* etc. are more likely to be terms candidates.

The results of contrastive word frequency search of MathWiki and BNC are given at the Fig. 13. The words like *trigonometric*, *polynomial*, *exponential*, *binomial(s)*, *hyperbola*, *ellipse*, *parabola*, *Dirichlet series*, *hypotenuse*, etc. are more likely to be terms candidates.

7.2 Linguistic analysis

The terms extracted represent simple words like *algebra*, *divisor*, *calculus*, *axiomatization*, *exponentiation*, etc. The compound derivative words like *x-intercepts*, *multi-valued*, *subinterval*, *semigroup*, *interpolant*, *permutation*, *double-angle*, *half-angle*, *holomorphic*, *single-valued*, etc. Abbreviations like *cos*, *sin*. The proper names terms like *Dirichlet series*, *Riemann series*, *Fourier series*, *Peano axioms*, etc. were extracted mostly from MathWiki corpus.

With respect to the part-of-speech, mostly nouns togheter with phrasals are most frequent. Terms are evaluated on the base of their low frequency or no occurence in the BNC. Thus, verbs typical for the mathematical precalculus texts like *triangularize*, *axiomatize*, and *subscripte* were not occured in BNC. Also, nouns like *hyperbola*, *trinomial*, *polynomial*, and adjectives like *trigonometric*, *multiplicative*, *non-*

Word list

Corpus: MathWiki

Reference corpus: preloaded/bnc

Page 1 Go

word	user/vstoykova/MathWiki Freq	Freq/mill	preloaded/bnc Freq	Freq/mill	Score
θ	137.0	1214.1	0	0.0	1215.1
trigonometric	141.0	1249.5	7.0	0.1	1177.1
π	131.0	1160.9	0	0.0	1161.9
polynomial	210.0	1861.0	90.0	0.8	1033.1
conic	88.0	779.8	6.0	0.1	741.2
cosine	80.0	708.9	36.0	0.3	537.5
exponential	86.0	762.1	102.0	0.9	399.7
codomain	43.0	381.1	0	0.0	382.1
binomial	44.0	389.9	44.0	0.4	280.8
hyperbola	36.0	319.0	21.0	0.2	269.6
tangent	53.0	469.7	100.0	0.9	248.8
calculus	51.0	451.9	108.0	1.0	230.8
ellipse	40.0	354.5	76.0	0.7	211.9
Frege	27.0	239.3	17.0	0.2	208.6
Cartesian	46.0	407.6	114.0	1.0	202.7
nonzero	28.0	248.1	32.0	0.3	193.8
trigonometry	24.0	212.7	20.0	0.2	181.3
conics	20.0	177.2	0	0.0	178.2
parabola	27.0	239.3	42.0	0.4	174.8
sines	21.0	186.1	9.0	0.1	173.2
logarithmic	29.0	257.0	60.0	0.5	168.1
Dirichlet	19.0	168.4	2.0	0.0	166.4
quadratic	30.0	265.9	78.0	0.7	157.4
Euclidean	24.0	212.7	47.0	0.4	150.6
hypotenuse	22.0	195.0	54.0	0.5	132.3
secant	14.0	124.1	2.0	0.0	122.9
subtraction	24.0	212.7	83.0	0.7	122.8
axiom	22.0	195.0	88.0	0.8	109.8

Page 1 Go
Next >

Figure 13: Contrasting word frequency lists for MathWiki and BNC corpora.

Collocation candidates

Page 1 Go

	Freq	T-score	MI	MI3
p/n zeta	6	2.435	7.444	12.614
p/n Dirichlet	5	2.187	5.518	10.162
p/n Frege	7	2.587	5.497	11.111
p/n Riemann	3	1.689	5.329	8.499
p/n Fourier	3	1.669	4.781	7.951
p/n Taylor	3	1.636	4.171	7.341
p/n definitions	3	1.636	4.171	7.341
p/n series	16	3.608	3.351	11.351
p/n plane	5	2.010	3.307	7.950

Figure 14: Collocations candidates of the concept *function* estimated from MathWiki corpus by using the $MI - score$ criterion.

negative, and, surprisingly, the term *precalculus* were not occurred in the BNC as well.

8 Defining highly specialized terminology

The former analysis uses extensively various semantically-oriented statistical search techniques to define the basic conceptual relations in the domain of precalculus with application to education. It, also, apply the techniques to extract the terminological relations in the same domain by comparing different statistical techniques.

However, the interpretation of the results was made mostly by using $T - score$ criterion, even the results are presented, also, for $MI - score$ and $MI^3 - score$ criteria. In further description, we are going to use the $MI - score$ criterion to evaluate the proper names terms.

As it was outlined in the previous section, the proper names terms were extracted mostly from MathWiki corpus. At the same time, the proper names terms represent, usually, a higly specialized terminology. Fig. 14 presents the collocations candidates of the concept *function* estimated from MathWiki corpus by using the $MI - score$ criterion. Alternatively, all collocated words were not occurred in the MathPre and MathWeb corpora which suggests that they refer to highly specialized concepts.

Corpus: MathWiki

Hits: 5 **(44.3 per million)**

usually meets the Dirichlet definition of	function	in his introductory course in calculus,
|Imaginary unit | (This	function	of x is the Dirichlet kernel.) Double-,
the Dirichlet kernel coincides with the	function	's nth-degree Fourier approximation. The
Riemann zeta function. Like the zeta	function	, Dirichlet series in general play an important
example, the Dirichlet series for the zeta	function	converges absolutely when Re s > 1, but

Corpus: MathWiki

Hits: 3 **(26.6 per million)**

Fourier claimed that every	function	had a Fourier series, something no mathematician
the Dirichlet kernel coincides with the	function	's nth-degree Fourier approximation. The
series is the Fourier series of a	function	. History of the theory of infinite series

Corpus: MathWiki

Hits: 3 **(26.6 per million)**

coefficients of a Taylor series of any rational	function	satisfy a linear recurrence relation, which
which can be found by setting the rational	function	equal to its Taylor series and collecting
operator, which acts on the Taylor series of a	function	. A matrix is a rectangular arrangement

Figure 15: Co-occurrences of the concept *function* related to collocates *Dirichlet*, *Fourier* and *Taylor*.

Corpus: **MathWiki**

Hits: **16** (141.8 per million)

, the standard series of the exponential	function	converges to a real number because
Fourier, for example, claimed that every	function	had a Fourier series , something no
a Taylor series of any rational	function	satisfy a linear recurrence relation, which
which can be found by setting the rational	function	equal to its Taylor series and collecting
When this series for the tangent	function	is expressed in a form in which the
When this series for the secant	function	is expressed in a form in which the
of characterizations of the exponential	function	; others involve series or differential
power series definition of the exponential	function	makes sense for square matrices (for which
the sum of this series is Riemann's zeta	function	. • A telescoping series converges
test. Series of functions Main article:	Function	series A series of real- or complex-valued
the Dirichlet series is the Riemann zeta	function	Like the zeta function, Dirichlet
Riemann zeta function Like the zeta	function	, Dirichlet series in general play an important
example, the Dirichlet series for the zeta	function	converges absolutely when Re s > 1, but
is the Fourier series of a	function	. History of the theory of infinite series
idea of an infinite series expansion of a	function	was conceived in India by Madhava of the
which acts on the Taylor series of a	function	. A matrix is a rectangular arrangement

Figure 16: Co-occurrences of the concept *function* related to its collocate *series*.

In fact, the term *zeta function* occurs on the first place followed by *Dirichlet*, *Frege*, *Riemann*, *Fourier*, etc. The concept *function* and the collocates *Dirichlet*, *Fourier* and *Taylor* are semantically related through the concept *series* as it is shown at Fig. 15.

Moreover, the semantic relation might be additionaly evaluated by generating the co-occurrences of the concept *function* related to its collocate *series*. The result is presented at Fig. 16 and it includes the collocates *Dirichlet*, *Fourier*, *Riemann* and *Taylor*.

Consequently, the concept *zeta function* and the concept *series* are related to *Dirichlet*, *Fourier* and *Taylor* expressing the specialized terminological relations of the concept *function*. The general expression of that relation is the representation of the *function* in *series*.

The above conceptual relations represent highly specialized mathematical knowledge and are studied

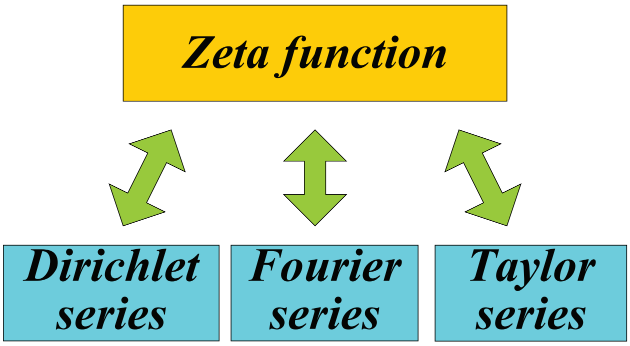

Figure 17: Conceptual semantic hierarchy of the concepts *function* related to its collocate *series* and the collocates *Dirichlet*, *Fourier* and *Taylor*.

at the university level of education. Their conceptual semantic hierarchy is presented at Fig. 17.

9 Building ontologies

As a very important area of AI, the knowledge representation techniques use extensively various approaches and techniques used in terminology. In general, building ontologies is, in fact, the creation of electronic dynamic terminological conceptual knowledge hierarchical on-line database, and use various techniques for semantically-oriented statistical search to define the co-occurrences and collocations.

The highly specialized proper names terms are usually developed as named entities in the ontologies framework. Thus, extracting basic mathematical terms of precalculus conceptual relations for building ontological hierarchy is already defined at Fig. 10.

10 Conclusion and future work

In our research, we have used three related web-based electronic corpora consisting of open-source mathematical texts about precalculus. The final results confirm that it is possible by using the statistically-based software incorporated in the SE by searching for keywords to define the basic mathematical concepts for teaching precalculus and to refine their conceptual hierarchy by searching for collocations and co-occurrences. Also, the methodology has been extended with application to terminology.

The BNC comparative results show very low frequency of the basic mathematical concepts for teaching precalculus. Surprisingly, we did not found there the term *precalculus* instead which suggests the idea that specialised corpora, even of relatively small size, are the most appropriate for specialize research instead standard ones, widely used for general linguistic research.

Further, we are going to continue our work by extending it for building a thesaurus-like conceptual hierarhy and by comparing the results for some more languages so to test the hypothesis for the language-independent nature of conceptual knowledge.

References:

[1] Bontchev, B. and Vassileva, D. (2011). Learning Objects Types Dependability on Styles of Learning. In N. Mastorakis, V. Mladenov, Z. Bojkovic, eds. *Recent Researches in Educational Technologies*, Corfu, Greece. 227-234.

[2] Church, K. and Hanks, P. (1991). Word Association Norms, Mutual Information and Lexicography. *Computational Linguistics* 16(1), 22-29.

[3] Hazewinkel, M. (2001). *Encyclopaedia of Mathematics.* Springer-Verlag.

[4] Kilgarriff, A. and Rundell, M. (2002). Lexical Profiling Software and its Lexicographic Applications: a Case Study. In *Proceedings from EURALEX 2002.* Copenhagen. 807-811.

[5] Kilgarriff, A., Rychly, P., Smrz, P., and Tugwell, D. (2004). The Sketch Engine. In *Proceedings from EURALEX 2004.* Lorient, France, 105–116.

[6] Kilgarriff, A., Reddy, S., Pomikalek, J. and Avinesh, P. (2010). A Corpus Factory for Many Languages. In N. Calzolari ed. *Proceedings of the LREC 2010.* Malta. 904-910.

[7] Lin, Dekang. (2002). Automatic Retrieval and Clustering of Similar Words. In *Proceedings of the COLING-ACL.* Montreal. 768-774.

[8] Mitkova, M. (2009). Achievement in Basic Math Courses Reflected Through the Gender Difference. In *Proceedings of the 4th International Conference on Research and Education in Mathematics (ICREM4)*, 76-81.

[9] Nitsolov, S. and Mitkova, M. (2011). Teaching Collisions - Methodological Suggestions. In N. Mastorakis, V. Mladenov, Z. Bojkovic, eds. *Recent Researches in Educational Technologies,* Corfu, Greece. 106-109.

[10] Oakes, M. (1998). *Statistics for Corpus Linguistics.* Edinburgh University Press.

[11] Precalculus. (2011). `http://en.wikipedia.org/wiki/Precalculus`

[12] Precalculus Tutorial. (2011). `http://jwbales.home.mindspring.com/precal/`

[13] Sparck Jones, K. (1986). *Synonymy and Semantic Classification.* Edinburgh University Press.

[14] Stoykova, V. (2002). Bulgarian noun – definite article in DATR. In D. Scott, ed. *Artificial Intelligence: Methodology, Systems, and Applications. Lecture Notes in Artificial Intelligence 2443*, Springer-Verlag, 152–161.

[15] Stoykova, V. (2009). Language-dependent or Language-independent e-Learning Mathematics. In *Proceedings, Part II, Tempus JEP 41110-2006, TEMIT*, 71–83.

[16] Stoykova, V. (2010). Representing Lexical Knowledge for Bulgarian Inflectional Morphology in DATR. In N. Mastorakis, V. Mladenov, Z. Bojkovic, and D. Simian, eds. *Latest Trends on Computers*, vol. 2, 612–616.

[17] Stoykova, V. and Mitkova, M. (2011). Defining Lexical Semantic Relationships for Terms of Precalculus Study. In N. Mastorakis, V. Mladenov, Z. Bojkovic, eds. *Recent Researches in Educational Technologies*, Corfu, Greece. 240-244.

[18] Topics in Precalculus. (2011). `http://www.themathpage.com`