

VI კონფერენცია

ბუნებრივ ენათა დამუშავება



Conference VI
Natural Language Processing

თბილისი 2008 Tbilisi

არნ. ჩიქობავას ენათმეცნიერების ინსტიტუტი

ბუნებრივ ენათა დამუშავება

2008 წლის 27-28 მაისი

კონფერენციის მასალები

მუშაობის გეგმა

სამშაბათი, 27 მაისი
11.00

ლ. ეზუგბაია, თ. უთურგაიძე, რ. სხირტლაძე – ზმნურ ფორმათა გენერაციის საკითხები

მ. ბერიძე, დ. ნადარაია – XX საუკუნის შიდამიგრაციების მონაცემთა ბაზა და საქართველოს ენობრივი სურათი

რ. ასათიანი – ენობრივი მასალის ბაზების სახით ორგანიზება და ანოტირებისათვის ღირებული პარამეტრების შერჩევა

გ. გოგიჩაიშვილი, ს. სერგევეი, ლ. ჩხაიძე – ბუნებრივი ქართული ენის ტექსტების რელევანტობის შეფასების ლინგვისტური მოდელი (სამოქალაქო სარჩელის ფორმულის მაგალითზე)

მ. სანადირაძე, დ. ქართველიშვილი, ლ. ჩხაიძე, ი. ჯანაშია, დ. ჯაში – ინტერნეტის ქართული სექტორის საძიებო სისტემა „უღვევის“ ავანპროექტი

ნ. სურმავა, მ. ბერიძე – კორპუსის ლინგვისტიკის ძირითადი ცნებანი და მათთვის ქართული შესატყვისების შერჩევის ცდა

ნ. ფუტყარაძე – საინტერნეტო დიალოგების (ფორუმების) ტერმინოლოგიისათვის (ქართული, რუსული და ინგლისური მასალის მიხედვით)

დ. ანფიმიადი – ვირტუალური სივრცის გავლენა და თანამედროვე ქართული პოეზიის ენის სპეციფიკა (რამდენიმე ქართული ლიტერატურული საიტის მაგალითზე)

დ ი ს კ უ ს ი ა

ოთხშაბათი, 28 მაისი

11.00

თ. გაშაკიძე, თ. ტეტელოშვილი, ნ. ჯორბენაძე – „ელექტრონული სასკოლო ორთოგრაფიული ლექსიკონის“ აგების თეორიული პრინციპებისათვის

ჯ. ანთიძე, ნ. გულუა – ქართული წინადადებების სრული კომპიუტერული მორფოლოგიური, სინტაქსური და ნაწილობრივ სემანტიკური ანალიზისა და ქართული ენიდან მანქანური თარგმანის შესახებ

გ. ჩიჩუა – მეტყველების გამოცნობის საცდელი პროგრამა

მ. ბარიხაშვილი – პროექტი „პერეთი ვირტუალურ სივრცეში“ (I ეტაპი – სამეცნიერო-ინფორმაციული ბაზის შესახებ)

კ. ფხაკაძე, ს. მასხარაშვილი, ლ. აბზიანიძე, ლ. გურასაშვილი, ნ. ფხაკაძე – ქართული ენისა და აზროვნების მათემატიკური თეორიის ზოგიერთი საკითხი

კ. ფხაკაძე, კ. გაბუნია, გ. ჩიჩუა, ა. მასხარაშვილი, ლ. აბზიანიძე, ნ. ვახანია, ნ. ფხაკაძე, ბ. ჩიქვინიძე, ლ. გურასაშვილი, ნ. ლაბაძე, მ. ბერი-აშვილი – ქართული ინტელექტუალური კომპიუტერული სისტემის კონსტრუირების მიზნები და ქართული ენის წინაშე მდგარი საფრთხეები

ქ. დათუკიშვილი – სახელთა მორფოლოგიური ანოტირების საკითხები ქართული ენის კორპუსისათვის

დ ი ს კ უ ს ი ა

კონსტანტინე ფხაკაძე, სანდრო მასხარაშვილი,

ლაშა აბოიანიძე, ლაშა გურასაშვილი, ნიკოლოზ ფხაკაძე

ი. ვეკუას სახ. გამოყენებითი მათემატიკის ინსტიტუტთან არსებული „ლოგიკისა და ენის გაერთიანებული ქართული ჯგუფი“ და „ქართული ენის ლოგიკისა და გამოთვლადის ახალგაზრდული ჯგუფი“
k_pkhakadze@caucasus.net

ქართული ენისა და აზროვნების მათემატიკური თეორიის ზოგიერთი საკითხი

ამა თუ იმ ბუნებრივი ენის ინტელექტუალური კომპიუტერული სისტემის კონსტრუირება ამ ბუნებრივი ენით მოცემული აზროვნებითი სისტემის მათემატიკური თეორიის შეუმუშავებლად ყოველად უპერსპექტივოა. ამასთან, სრულიად ნათელია, რომ ქართული ენისა და ქართული ენით მოცემული აზროვნების მათემატიკური თეორია ამომწურავი სიზუსტით უნდა ასახავდეს ამ ენისა და ამ ენით მოცემული აზროვნების თავისებურებებს და რომ იგი ვერც ერთი სხვა ენისა და აზროვნების მათემატიკური თეორიით ვერ ჩანაცვლდება.

ამგვარად, ზემოხსენებული მიზნით წარმოებული ჩვენი კვლევები თავიდანვე ორიენტირებული იყო არა ამავე მიზნებით დიდი ხნის წინ დაწყებული და დღესაც აქტიურად მიმდინარე საზღვარგარეთული კვლევების ანალოგიურობაზე, არამედ ქართული ენისა და აზროვნების ორგანულად დამახასიათებელი ბუნებრივი წესრიგის მაქსიმალურად ზუსტ და სრულ მათემატიკურ ასახვაზე.

ქართული ენისა და აზროვნების მათემატიკური თეორიის საფუძვლები შემუშავდა 2001-2007 წლებში თსუ სახელმწიფო მიზნობრივი პროგრამის „კომპიუტერის სრული პროგრამულ-მომსახურებითი მოქცევა ბუნებრივ ქართულ ენობრივ გარემოში“ კვლევითი ამოცანების გათვალისწინებით. კერძოდ, ამ კვლევებით შემუშავდა ქართული ენის პირდაპირი ანუ სიტყვასიტყვითი აღწერის მეთოდი და თითქმის დადასტურდა ენის ამგვარი აღწერის შესაძლებლობა. აქედან გამომდინარე, გაკეთდა უმნიშვნელოვანესი განაცხადი ქართული ენისა და მათემატიკური ენის ზოგადი ერთეობობრიობის თაობაზე. გარდა ამისა, დადგინდა ქართული წერითი ენის მარტივი თხრობითი წინადადების მათემატიკური ბუნება და თითქმის სრული მოცულობით გადაწყდა ამგვარი წინადადებების ქართულ აზრით ენაზე ავტომატური დაყვანის პრობლემა. ამასთან, ქართული აზრით ენის ბირთვული ნაწილი გაიწერა ფორმალურად განვითარებადი მათემატიკური თეორიის სახით და დადგინდა ქართული აზრით ენისა და ქართული წერითი და ზეპირი ენების შინაარსობრივი ურთიერთკავშირების მარეალიზებული ფორმალური, ლოგიკური და სიმრავლური მექანიზმები.

მოხსენებისას წარმოდგენილი იქნება ქართული ენისა და აზროვნების მათემატიკური თეორიის საფუძვლების შემდგომი დაზუსტებისა და გაღრმავების მიზნით წარმოებული კვლევების შედეგები. კერძოდ: გაკეთდება ქართული ენის ზოგიერთი რთული წინადადების მათემატიკური

კური ანალიზი, რის საფუძველზედაც ზოგადად მოიხაზება ამ ტიპის წინადადებების ქართულ აზრით ენაზე ავტომატურად დაყვანისა და ამ ენაში ავტომატურად დამუშავების მეთოდები; პროფ. შ. ფხაკაძის აღნიშვნათა თეორიის მიხედვით მოხდება ქართულ ენაში არსებული ზოგიერთი შემამოკლებელი სიტყვისა და შემოკლებული ენობრივი გამოხატულების კლასიფიცირება; წარმოდგენილი იქნება ქართული წერიითი ენის მართლმწერი და მათემატიკურ ენაზე მართლმთარგმნელი გაფართოებადი პროგრამული სისტემის ექსპერიმენტული ვარიანტი, რომელიც ჯგუფში უკვე შემუშავებული მათემატიკური მეთოდებით არის აგებული.

კონსტანტინე ფხაკაძე, კახა გაბუნია, გიორგი ჩიჩუა,
ალექსანდრე მასხარაშვილი, ლაშა აბზიანიძე, ნიკოლოზ ვახანია,
ნიკოლოზ ფხაკაძე, ბესარიონ ჩიქვინიძე, ლაშა გურასაშვილი,
ნინო ლაბაძე, მარი ბერიაშვილი

ი. ვეკუას სახ. გამოყენებითი მათემატიკის ინსტიტუტთან არსებული „ლოგიკისა და ენის გაერთიანებული ქართული ჯგუფი“ და „ქართული ენის ლოგიკისა და გამოთვლელის ახალგაზრდული ჯგუფი“
k_pkhakadze@caucasus.net

ქართული ინტელექტუალური კომპიუტერული სისტემის კონსტრუირების მიზნები და ქართული ენის წინაშე მდგარი საფრთხეები

ბუნებრივენობრივი ინტელექტუალური კომპიუტერული სისტემების კონსტრუირების მიზნით უკვე ორმოცდაათ წელზე მეტია მიმდინარეობს კვლევითი პროცესები. ამ პროცესების მიზანია ისეთი კომპიუტერის კონსტრუირება, რომელთანაც მომხმარებელს საშუალება ექნება იქონიოს სრული და თავისუფალი ბუნებრივენობრივი ურთიერთობა ნებისმიერი ინტელექტუალური თუ იმპერატიული ამოცანის გადაწყვეტისას.

ასეთი ქართული კომპიუტერის არსებობის შემთხვევაში ქართველ მომხმარებელს საშუალება ექნება ქართულად მისცეს ბრძანება კომპიუტერს, რომელიც გაიგებს და შეასრულებს ამ ბრძანებას.

ასე მაგალითად, მომხმარებელს საშუალება ექნება ასეთ კომპიუტერს ამოსახსნელად მისცეს ზოგადი უნარების გამოცდებისას გამოყენებული ვერბალური თუ მათემატიკური სახის ამოცანები და კომპიუტერი საკმაოდ ზუსტ, სრულ და ამომწურავ პასუხებს გასცემს მათ. კომპიუტერს არა მარტო ტესტურ რეჟიმში ექნება ასეთ ამოცანებზე პასუხის გაცემის საშუალება, არამედ იგი მის მიერ გაკეთებული არჩევანის მართებულობის ლოგიკურ დასაბუთებასაც შეძლებს.

კომპიუტერს შესაძლებლობა ექნება განახორციელოს ენობრივი ტექსტის დანაწევრება და გააზრება მორფოლოგიური, სინტაქსური,

ლოგიკური და სემანტიკური ანალიზისა და ზოგადი ენობრივი აზროვნებისთვის აუცილებელი სპეციალური მათემატიკური და გეომეტრიული ცოდნის საფუძველზე, რომელიც სისტემურად იქნება უზრუნველყოფილი მასში.

მოხსენებისას უფრო ვრცლად შევეხებით ამ პრობლემურ საკითხებს. ჩვენი აზრით, ქართველი ენათმეცნიერების, მათემატიკოსებისა და კომპიუტერულ მეცნიერებათა სპეციალისტების სახელმწიფოებრივად ორგანიზებული მიზნობრივი მუშაობის შემთხვევაში ქართველი ერი შეძლებს ქართული ენა ააშოროს დღეს ინტენსიურად მიმდინარე ბუნებრივენობრივი ტექნოლოგიური პროცესებით განპირობებულ საფრთხეებს.

გიორგი ჩიჩუა

ი. ვეკუას სახ. გამოყენებითი მათემატიკის ინსტიტუტთან არსებული „ლოგიკისა და ენის გაერთიანებული ქართული ჯგუფი“ და „ქართული ენის ლოგიკისა და გამოთვლადების ახალგაზრდული ჯგუფი“
gchichua@yahoo.com

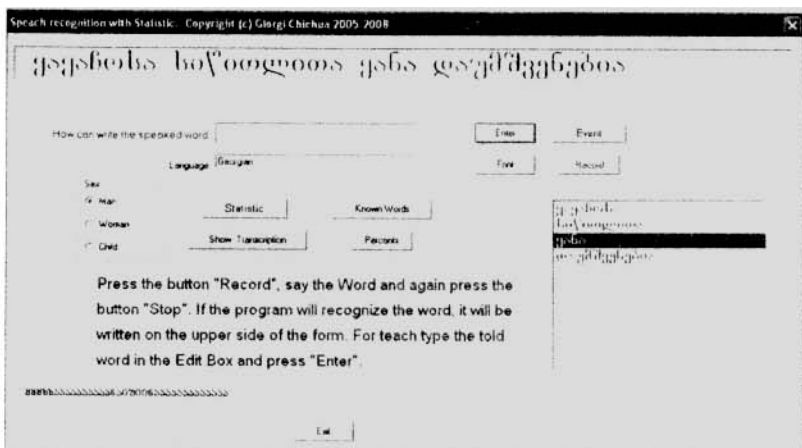
მეტყველების გამოცნობის საცდელი პროგრამა

მსოფლიოს მრავალ ქვეყანაში ინტენსიურად მუშაობენ ბუნებრივი ენის კომპიუტერული დამუშავების პრობლემატიკაზე. კერძოდ მეტყველების სინთეზზე, გამოცნობაზე, გაგებასა და თარგმნაზე. ჩვენც თუ ქართული ენისთვის არ შევქმენით ანალოგიური სისტემები, ქართულ ენას პერსპექტივაში ემუქრება დეკლუტურიზაცია. ადამიანები (ამ შემთხვევაში საქართველოს მოსახლეობა) თავის ყოველდღიურ მოღვაწეობას გააგრძელებენ იმ ენაზე, რომლისთვისაც შემუშავებულია კომპიუტერთან ურთიერთობის სისტემები. კერძოდ ინგლისური ან რუსული ენა გამოდგენის ქართულს.

აღნიშნულიდან გამომდინარე, ფრიად აქტუალურია სამუშაო, რომელსაც ახორციელებს ჩვენი ჯგუფი „ენისა და ლოგიკის გაერთიანებული ქართული ჯგუფი“. ერთ-ერთი ამოცანა, რომლის გადაწყვეტასაც ჩვენ ვცდილობთ, არის ქართული მეტყველების გამოცნობის პროგრამის შექმნა, რომლის საშუალებითაც მომხმარებელი შეძლებს Word-ის დოკუმენტში ქართული ტექსტის ხმის საშუალებით შეტანას.

ამ ეტაპზე შექმნილი გვაქვს პროგრამის საცდელი ვერსია, რომელიც გამოიცნობს ქართულ მეტყველებას 97%-ზე მეტი სიზუსტით 300-ზე მეტსიტყვიანი ლექსიკონიდან. სისტემას შეიძლება დაემატოს ახალი სიტყვები.

პროგრამის ინტერფეისი შემდეგნაირია:



ნახატზე დაფიქსირებულია მომენტი, როდესაც პროგრამამ გამოიცნო ფრაზა „ქაქაქოსა სიწითლითა ყანა დაუმშვენებია“. გამოცნობილი ტექსტი ზედა რედაქტირების ველში შეიძლება მოინიშნოს და გადატანილი იქნეს რაიმე ტექსტურ რედაქტორში. თუ პროგრამამ რომელიმე სიტყვა არასწორად გამოიცნო, შეიძლება ეს სიტყვა მოვნიშნოთ სიაში და პროგრამას ვასწავლოთ მისი სწორი დაწერილობა. ამ შემთხვევაში მონიშნულია სიტყვა „ყანა“.

ხმის გამოცნობისთვის ვიყენებთ ხმის სიგნალის ოსცილოგრამაზე გადატანილ ბადეს და სიგნალის გრაფიკის ამ ბადესთან გადაკვეთის წერტილების ანალიზით ვიღებთ სპექტრული სურათების მსგავს სურათებს, რომლებსაც შემდგომში ვიყენებთ გამოცნობის ევრიტიკულ მეთოდიკაში საბაზო მონაცემებად.

G. Chichua

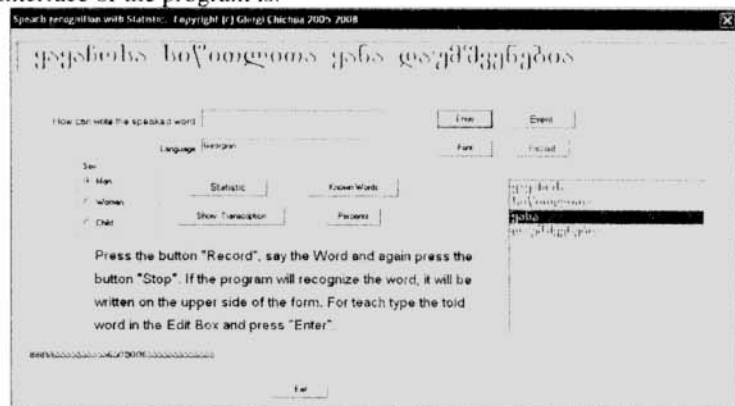
Trial Program of Speech Recognition

There are many countries in the world intensively working on a problem of computer processing of natural languages. Particularly on speech synthesis, recognition, understanding and translation. If we don't create analogous systems for Georgian language, it will be threatened with decultivation in perspective. People (in this case Georgian people) will continue their everyday public work in the language, for which is developed systems for relationship with computers. For example English and Russian Languages will excel Georgian language.

It is followed from mentioned above that it is highly actual work which is being realized by our group "United Georgian Group of Language and Logic". One of the problems we are trying to solve is creation of Georgian speech recognition program, with the help of which customer will be able to type text in a word document by voice dictation.

At the present time we have created a trial version of the program, which recognizes Georgian speech at accuracy more than 97% with Dictionary about 300 words. New words can be added to the system.

Interface of the program is:



The image shows a moment when program has recognized a phrase "The Poppy has adorned the cornfield with its redness". Recognized text in edit field above can be selected and pasted in some text editor. If program has incorrectly recognized some word, we can select this word in the list and teach its correct spelling. At this moment "cornfield" is selected.

We use stretched net on the oscilogram of signal for speech recognition and with analysis of intersect points of signal graphic and net we get images like spectral images, which we use as basic data in euristic methods of recognition.

predicates (i.e. without personal verb forms), just nouns, adjectives and adverbs united into special structure.

The problem is solved by means of parallel linguistic analysis of both texts, as a result of which formalized computer images of their contents are formed, allowing to strictly define the level of relevancy, which is the direct answer to the problem in question. This analysis consists of several stages:

1. Morphologic analysis of correct word-forms of the given syntagmae.
2. Location of a syntagmae actor - the forming element of its syntax structure, to which all the other text words are syntactically subordinate.
3. Location of the first level objects to which the action of the found syntagmae actor is directly performed.
4. Location of the second level objects, directly subordinate to the first level objects, which usually are adverbs or adverbial constructs, determining conditions or circumstances of actor action application.
5. Location of the third level objects, which represent the delicate syntax structure of attribute constructs of the previously indicated objects and related adverbs.
6. Determination of the principal causative relation of syntagmae directed from the actor to the first level object.
7. Identification of subordinate causative and qualitative relations between the first level objects and subordinate higher level objects.
8. Devising schemes for causative-qualitative relations between syntax elements of the two given texts allows strict formal description of the problem of determining their relevancy.

K. Pkhakadze, K. Gabunia, G. Chichua, A. Maskharashvili, L. Abzianidze, N. Vakhania, N. Pkhakadze, B. Chikvinidze, L. Gurasashvili, N. Labadze, M. Beriashvili

The Aims of Constructing Georgian Intellectual Computer System and Cultural Perspectives of Georgian Language

More than thirty years, in many countries, the research processes in order to construct intellectual computer systems in their native languages are already going on. The aim of those processes is to construct such computer with which the users, in the case of any kind intellectual or imperative necessity, will have possibility to make free and complete communication based only on their native language knowledge.

It means that in the case of "existence" such Georgian intellectual computer system the users will be able to say their commands to the computer in Georgian and the computer will understand and execute it. Also, the abilities of this computer will be so capable and, accordingly, the spectrum of such commands will be so wide that

this computer system, without any overstatement, can be regarded as the active intellectual partner of its users in almost every kind of societal necessity.

For example, user will be able to give a task used at the general skills exams to such computer to solve and the answer of this computer, in many cases, will be more complete than the one of many examiners. We mean that this computer will be able to solve such task not only in tests mode, but also it will be able to give Georgian verbal prove of its decision in textual or in speech form. This will be performed with such type softwaring of it, by which the computer will be able to disjoin, analyze and understand Georgian speech and writing. This will be realized by time-parallel morphological, syntactical, logical and semantical parsing and by the basis of general mathematical knowledge.

Also, only such computer, software core of which realized its general Georgian lingual, logical and mathematical knowledge, will make reachable to organize special Georgian intellectual computer expert systems of any different kinds.

It is clear, that the future streaming of general world-wide cultural processes is not possible to consider without active and fundamental participation of such computer systems. And this, on the other hand, makes clear, that in the case of not "existence" such general Georgian intellectual computer system Georgian people will lose ability to take part in the future planned cultural processes by means of Georgian language.

During our report we will describe this very problematic points in details and, also, it will be proved that in the case of the state priority organization of joint acting of Georgian linguists, mathematicians and specialists of computer sciences, and only in this case, the Georgian language, i.e. Georgian Nation, is able to overcome the threats implied by contemporary technological processes.

**K.. Pkhakadze, A.. Maskharashvili, L. Abzianidze,
L. Gurasashvili, N. Pkhakadze**

Some Aspects of the Mathematical Theory of the Georgian Language and Thinking

It is clear that to construct the intellectual computer system for a natural language system without mathematical theory of this language and thinking is impossible. Besides of this, it is clear that mathematical theory of Georgian language and thinking must represent direct and detailed description of the natural specifics of this one. Also, it is clear, that this Georgian theory can not be replaced by the mathematical theory of any other natural language and thinking.

According above, our researches for the elaboration mathematical theory of the Georgian language and thinking were, are and will be pointed on the mentioned natural specifics of the Georgian language and thinking. Moreover: According ours view our scientific duty is to recognize the rules of our natural language knowledge, which are in us mainly on the under cognition level. At the same time: According to

our view any natural language knowledge is based on that general logical and mathematical knowledge, which has a universal nature and which is in humans as the instinctually or automatically organized deepest knowledge.

On the basis of above shortly presented points of view the fundamental aspects of the mathematical theory of the Georgian language and thinking were worked out in 2001 - 2007. These researches were aimed on scientific problems of TSU State Priority Program "Complete Programming Inclusion of a Computer in Georgian Natural Language System".

Particularly: There is elaborated the method, which is called as natural, or as direct formal-logical description of Georgian language; There was proposed the very important statement: The statement declares that the types of Georgian and Mathematical languages, in generally sense, coincide to each other; There is studied the mathematical nature of a simple declarative sentence of Georgian written and oral languages and the problem of automatic reduction of such sentences into Georgian thought language has been almost entirely solved. In other words, it means that there is almost entirely solved the very important problem of automatic translation of simple declarative sentences of Georgian written and oral languages into mathematical language; There is alphabetically described (in mathematical sense) a core part of Georgian thought language; There is studied formal, logical and set-theoretic aspects of the core part of Georgian written and oral languages, which are mainly coincides with same aspects of the core part of Georgian thought language, which is understood as the formally developing mathematical theory.

At the lecture we will represent the results of researches which are made with the purpose of further wider and deeper investigation of mathematical theory of Georgian language and thinking.

More precisely: There will be lectured the mathematical methods of analyses of some complex sentences of Georgian language. Consequently there will be discussed methods of automatic reduction of these types of sentences into Georgian thought language; There will be presented the first short classification of abbreviated or abridge, or shorting words and expressions of Georgian language. This classification is made according to Notation Theory by Prof. Sh. Phkhakadze; There will be presented natural set-diagram approaches for Georgian language. In addition to these, there will be presented the experimental version of parsing and translator computer system for Georgian written language. The system, which parses simple Georgian sentences and translates it into the mathematical language, is equipped with the option that allows expand its types and words vocabulary.