

PART-18

Washington University in St. Louis
Danforth Campus



Central Eurasian Studies Society

INTERNATIONAL SCIENTIFIC-ONLINE

CONFERENCE ON INNOVATION IN THE MODERN EDUCATION SYSTEM

WASHINGTON

2021



Central Eurasian Studies Society



INNOVATION IN THE MODERN EDUCATION SYSTEM

Part 18

MAY 2022

COLLECTIONS OF SCIENTIFIC WORKS

Washington, USA
25th MAY 2022

СОВРЕМЕННАЯ КОНЦЕПЦИЯ ПРОФЕССИОНАЛЬНОГО ОБРАЗОВАНИЯ В СИСТЕМЕ ВЫСШИХ СПЕЦИАЛЬНЫХ ОБРАЗОВАТЕЛЬНЫХ УЧРЕЖДЕНИЙ	
Abdullaeva Nasibaxon Abdusalomovna TALABALARNING KASBIY KOMPOTENTLIGI	747
Toshmirzayev Qodirjon Odiljonovich Toshmirzayeva Dilshoda Odiljonovna TILSHUNOSLIK NAZARIYASI VA AMALIYOTI MASALALARI	752
Ким Светлана Геннадьевна ИННОВАЦИОННЫЕ ФОРМЫ И МЕТОДЫ ОБУЧЕНИЯ - КАК ОДИН ИЗ СПОСОБОВ ПОДГОТОВКИ УЧАЩИХСЯ К ДЕЙСТВИЯМ ПРИ СИЛЬНЫХ ЗЕМЛЕТРЯСЕНИЯХ	755
Турапов Р.А ВЛИЯНИЕ ВНЕШНИХ ФАКТОРОВ НА ИНВЕСТИЦИОННУЮ ПРИВЛЕКАТЕЛЬНОСТЬ ОБЪЕКТА	759
Turaeva Sevara DIGITALIZATION AND THE DIGITAL ECONOMY AND THE CHANGES THEY INTRODUCE INTO OUR LIVES	764
Гаюпов Ж Бородина М.Р ТУРКЕСТАНСКИЙ МОДЕРН И ХУДОЖЕСТВЕННАЯ КОВКА В КОКАНДЕ	767
Abbasov Farukh Feruzovich TASKS OF EDUCATING YOUNG PEOPLE IN THE SPIRIT OF MILITARY PATRIOTISM	780
Tairova Malika Akramovna HISTORY OF RUSSIAN LANGUAGE DEVELOPMENT AND ITS IMPORTANCE IN THE WORLD	785
Isabayeva Asida Yusufjonovna MAKTABGACHA YOSHDAGI BOLALARDA NUTQNI TINGLASH, FIKRLASH VA MULOQOTNI RIVOJLANTIRISH UCHUN MASHG'ULOT DAVOMIDA O'TKAZILISHI MUMKIN BO'LGAN O'YINLAR.	790
Мақсад Исабаев МАҲАЛЛИЙ БОШҚАРУВДАГИ НОМАРКАЗЛАШУВ ЖАРАЁНИДА ҲОКИМИЯТ ОРГАНЛАРИ ВА МАҲАЛЛИЙ ЎЗИНИ ЎЗИ БОШҚАРИШ КОНЦЕПЦИЯСИ	794
Хайдаров Улуғбек Панжиевич ИПЛАРНИНГ ФИЗИК-МЕХАНИК ХОССАЛАРИНИНГ ЎЗГАРИШИ	805
Makhamatkarimov Jakhongir Dilmurod og'li Uzakova Z. INCLUSIVE EDUCATION IN THE REPUBLIC OF UZBEKISTAN DEVELOPMENT PROSPECTS	809
Naimova Dildora Kahramonovna MORPHOLOGICAL ANALYSIS OF THE TEXT	812
Булекбаева Екатерина Александровна Хужанова Озода Таджиевна АНАЛИЗ ТВОРЧЕСТВА УЗБЕКСКОГО НАРОДНОГО ПИСАТЕЛЯ ШУКУРА ХОЛМИРЗАЕВА	815
Наврузов Дилшод Примкулович Абдухамидов Сардор Кахарбоевич ДВУХШАГОВАЯ НЕЯВНАЯ СХЕМА ПИСМЕНА-РИКФОРДА ДЛЯ РЕШЕНИЯ УРАВНЕНИЕ ЛАПЛАСА.	818
Nafasov Mirzomurod Muxamadovich O'ZBEKISTON SHAROITIDA TA'LIMIY MOBIL ILOVALARDAN FOYDALANISH	821

MORPHOLOGICAL ANALYSIS OF THE TEXT

<https://doi.org/10.5281/zenodo.6582241>

Naimova Dildora Kahramonovna

Uzbekistan State University of World Languages

Abstract: *One of the foundations of a language is its vocabulary. However, what does a dictionary contain? Since the term "word" is too semantic for using it in a strict scientific text. The article deals with several concepts of the use of words in the text and morphological analysis.*

Key words: *Punctuation, scientific, concept, grammar, morphologic analysis, lexeme, phraseology, context.*

Introduction

One of the basics of a language is its vocabulary. But what does a dictionary contain? Since the term "word" is too ambiguous to be used in a strict scientific text, let's introduce a few concepts. If we extract all substrings from the text that do not contain separators (spaces, some punctuation marks, etc.), we will get a lot of tokens. For example, the token will be the words entrance or under (since they can occur in the text on their own), but the substring under will not be (unless, of course, it is written in the text in this form). It is believed that for each token there is its initial (or normal) form (also called a lemma) From this initial form, all other forms of the word are created by inflection, that is, some changes in this initial form.

The formation of new words or their forms occurs at the level of combining morphs - the minimum meaningful units of the language. Morphs are divided into root (word root) and service: prefix (prefix), suffix, inflection (ending), postfix. The carrier of the main meaning of the word is the root, and the service ones, in the general case, give an additional meaning. The division of a word into morphs is called morphemic parsing. Some service morphs (for example, prefixes and suffixes) are responsible for the formation of new words, others (for example, endings) are responsible for the formation of word forms. Changing the form of a word is tied to a set of grammatical parameters (tags): part of speech, gender, number, case, possessiveness, and so on. By a word form, we mean a group (tuple) consisting of a token, an initial form associated with it, and a set of grammatical parameters¹. For example, the word form is the set {feline, feline, adj. husband. genus, unit h., genus. n}, which contains the string

"feline", associated with the initial form "feline", and which is characterized by the following parameters: adjective in the masculine, singular and genitive case. By a lexeme we mean the set of all word forms associated with a given initial form.

Based on these definitions, we can say that the task of morphological analysis (lemmatization) is to find the word of the form in the dictionary by its token. The task of morphological synthesis is just the opposite: according to the word form, it is necessary to return its token.

Or more formally:

- morphological analysis is the derivation of a lemma or basis (pseudo-foundation) of a given token, and, if necessary, morphological parameters;
- morphological synthesis is the generation of the desired word form of a meaning or its entire paradigm according to the normal form (or base) and morphological characteristics.

By word usage, we mean the occurrence of a word form in a text. Depending on the context, word usage can be understood as either only a word form string, or a word form as a set. For example, the phrase "Slanting oblique mowing oblique after sandy spit" contains 8 word usages, 7 word forms, 6 lexemes and 4 unique tokens. This becomes obvious if you write word forms instead of strings.

Such a form is called a list form and is convenient in a number of cases: it is more convenient to read individual parameters, the very presence of a parameter with this name can be an indicator, the names of the parameter values can coincide, the parameters can follow in an arbitrary order, the number of parameters is variable. In some situations, positional notation is more convenient, when the first letter denotes a part of speech, followed by a fixed number of letters that specify parameter values. The number of parameters is given by the part of speech; the sequence of parameters is fixed. Such a record is convenient in cases where the list of parameters will not change. It takes up less space, but it requires additional information about the location of the parameters.

Before conducting a morphological analysis, it is necessary to isolate individual words from the text. In this regard, a graphematic analysis subsystem is sometimes supplied with the morphological analysis system. The input stream of characters is divided into tokens of several classes: alphabetic sequences, numbers, alphanumeric complexes, punctuation, separators, hieroglyphs. At the same time, each class of tokens has its own set of tags, in particular, for words it can be language (Cyrillic or Latin) and register.

In connection with the presence of such phenomena in the language, it is necessary to introduce a module for the analysis of non-dictionary words into the system of morphological analysis. It is usually implemented using a set of heuristics such as prefix stripping, ending analogy, and rules for hyphenated words. Another important function is the removal of morphological homonymy. In different systems, two different approaches are implemented to solve this problem: contextual and contextless removal. Context-free removal is performed based on the calculation of statistics for the marked corpus, and contextual removal is performed using a classifier configured using one of the machine learning methods.

LITERATURE:

1. Chomsky, N. Syntactic Structures. The Hague: Mouton, 1957.
2. Feldman R., Sanger J. (ed.). The text mining handbook: advanced approaches in analyzing unstructured data. — Cambridge University Press, 2007.
3. Fillmore C. J. The Case for Case. In: Bach and Harms (Ed.): Universals in Linguistic Theory. New York: Holt, Rinehart, and Winston, 1968.
4. Grishman R., Information Extraction. In: The Handbook of Computational Linguistics and Natural Language Processing. A. Clark, C. Fox, and S. Lappin (Eds), WileyBlackwell, 2010, pp. 515-530.
5. Harabagiu, S., Moldovan D. Question Answering. In: The Oxford Handbook of Computational Linguistics. Mitkov R. (ed.). Oxford University Press, 2003, p. 560-582