

ЦИФРОВА ПИСМЕНОСТ И МЕТОД ЗА КОМУНИКАЦИЯ НА МАЙЧИН ЕЗИК

(Патент BG 63704 B1)



Койчо Митев
Михаела Шимкович
Георги Митев

Колкото по-невероятна изглежда една идея, толкова е по-възможно, тя да се окаже истина.

Закон на Мърфи

"....Джон Атанасов живо се интересуваше от най-важното средство за предаване на информация - живия човешки език и неговия писмен еквивалент. Писмените езици, по отношение на говоримите, са някакво символно представяне на последните, което не е еднозначно определено и следователно може да има различно качество. По повод на неочакваното от него награждаване с орден "Кирил и Методий" - първа степен, Джон Атанасов показва рядка за един американец осведоменост за делото на братята Кирил и Методий. Това се дължеше на неговия дългогодишен интерес към различните видове писменост. Той често се оплакваше от високия процент неграмотност в САЩ, която обясняваше с несъвършенството на писмения английски език, като признаваше Кирилицата за много по-удачна.

Това е мотивирало у него желанието да сътвори нова писменост, която да е напълно фонетична и да е пригодна както за хората, така и за машините. Тази негова мечта остана непостигната, макар да се стремеше упорито към нея до края на живота си".

Академик Благовест Сендов -
"ВЕКОВЕН ЮБИЛЕЙ НА ДЖОН АТАНАСОВ "

<http://www.aba.government.bg/bg/BGpoSveta/Kariera/021203.html>

Цел на изобретението:

Да се предложи нов, непознат до сега инструмент за комуникация на майчин език, посредством компютърни устройства.

Задача на изобретението:

- Да се състави цифрова писменост за звуковата и писаната реч на всички езици, които използват хората по света;
- Цифровата писменост да съдържа информация за всички фонетични и граматически характеристики на речта, които да бъдат достатъчни за точна комуникация на майчин език;
- Цифровите кодове на речта да бъдат свързани с база данни от междуезикови речници така, че комуникацията да бъде възможна в реално време;
- Цифровата писменост да позволи върху електронен носител - електронен чип или процесор на РС да бъде записана информацията за речника на множество езици с едни и същи знаци (цифрите от 0 до 9), които след превръщането им в бинарни кодове да са регенеруеми;
- Цифровата писменост да позволява извършването на морфологичен и синтактичен анализ на изреченията от превеждания текст на всеки произволен език, както и възпроизвеждането на превода в звук или текст с граматиката и словоредата на съответния език;
- Цифровата писменост да позволява комуникация на майчин език от вида: звук-към-звук (по телефон или микрофон), текст-към-текст, звук-към-текст и обратно за всички езици и диалекти.

Това изобретение е изградено върху известни научни факти.

Първи научен факт:

В резултат на своето историческо развитие почти всички народи са създали или приели писмени знаци, с които се записва човешката реч. Различните писмености са несъвместими. Те са част от световното културно наследство и се развиват по свой индивидуален път. Днес съществуват около 6800 езика по света, като прогнозата на учените е, че в края на това столетие половината от тях ще изчезнат. Ефектът от този процес вероятно е съпоставим с ефекта от изчезването на животински и растителни видове.

Втори научен факт:

Единствените общи писмени знаци във всички езици са цифрите (0-9) и знаците за пунктуация.

Трети научен факт:

Клавиатурите на компютрите по света са практически несъвместими, като единствените общи клавиши върху тях са за цифрите и знаците за пунктуация.

Четвърти научен факт:

Във всички езици по света граматиката съдържа едни и същи общи компоненти: фонетика, морфология, синтаксис, лексика, семантика.

Пети научен факт:

Математиката е универсален инструмент за анализ на механизмите, по които се движи природата. Тя може да бъде и инструмент за предаване на езикова информация (например бинарните кодове за запис и трансфер на

Информационни единици, които използва изобретателят на първия компютър Джон Атанасов в ABC computer).

Шести научен факт:

Съставните единици на всяка произволна човешка реч са: звук, фон, алофон, фонема, морфема, дума, словосъчетание, изречение, текст. (Не бива да се смесва с понятието лексикална единица).

Седми научен факт:

Говорният апарат на хомо сапиенс е един и същ за всички раси и не зависи от етническия произход.



[Http://artefact.lib.ru/languages/russian/1-32.html](http://artefact.lib.ru/languages/russian/1-32.html)

Осми научен факт:

Броят на фонемите, които използва хомо сапиенс в звуковата реч е двуцифрено число. Съществуват технологии, които позволяват да се различат отделните фонемите една от друга чрез прилагането на Диференциален признак (артикулярен или акустически признак на фонемите) и други средства от фонетиката.

Девети научен факт:

Подобно на пръстовите отпечатъци човешкият глас е уникален за всеки индивид. Ако бъде съставен списък от ограничен брой думи, които предварително са подбрани така, че да съдържат всички възможни позиции на фонемите от речта, то след еднократен запис на звук от конкретна човешка реч може да бъде създадена технология, която да разпознава тези фонемите.

ФОНЕТИЧНИ ЦИФРОВИ КОДОВЕ

PH code (Phonetic code)

Задачата на този код е да се идентифицират фонемите от разпознат глас с едни и същи цифрови кодове, които да бъдат приложими за всички езици. Независимо от знаците, с които се записват думите в различните писмености, звуците от речта имат общи характеристики. Структурата на цифровите кодове е както следва:

- Първите две цифри показват пореден номер на фонемите.
- Третата цифра показва дължина, продължителност на звука (0 нормална продължителност на звука, 1 дълга гласна или сдвоена съгласна).
- Четвъртата цифра показва наличието на признака „мекост“ при съгласните звуци (0 нормален звук, 1 мека съгласна), а при гласните звуци ударение (0 неударена гласна, 1 гласна в сричка под ударение).
- Пета и следващи цифри при необходимост биха могли да диференцират други характеристики на фонемите.

Пример:

Гласни звуци

A 0100 **Á 0110** >CZ: nemáе немааме;

E 0200 **É 0211** >SK: pekné пекнее;

I (I, Y) 0300 **Í, Ý 0310** >EN: easy ийзи;

O 0400 **Ó 0410**

U (U) 0500 **Ú 0510**

Ъ 0600 >EN: camera /'kæmrə/; open /'əʊpən/;

Ä 0700 >SK: mäso, EN: magazine;

Λ 0800 >EN: puff / pʌf /, blood / blʌd /;

Дифтонги (Diphthongs)

eɪ - 0900 >EN: day, came, made examination;

aɪ - 1000 >EN: time, high, design;

ɔɪ - 1100 >EN: boy, noise, joy, employ;

əʊ - 1201 >EN: shoulder гласна в ударена срочка;

1200 >EN: hypothetical /,haɪpəʊ'tetɪk/ - гласна в неударена срочка;

aʊ - 1301 >EN: house, town, sound ударени гласни;

ɪə - 1400 >EN: frontier, overseer неударени гласни;

1401 >EN: here, severe, engineer ударени гласни;

eə - 1501 >EN: pair, there, chair, wear ударени гласни;

ʊə - 1600 >EN: contour, manicure неударени гласни;

1601 >EN: sure, tour, moor ударени гласни;

yo (ô, uo) - 1700 >SK: rôvabný пуовабний (чаровен);

yo (iu) - 1800

ya (ja) - 1900

ye (je) - 1900 >CZ: medvěd - мечка;

y - 2000

.....

..... други до пореден номер 29.

Съгласни звуци

Б	(B)	-	3000	
В	(V, W)	-	3100	
Г	(G)	-	3200	
Д	(D)	-	3300	Ǿ- 3301 >SK: medved`
Ж	(J, Ź)	-	3400	
З	(Z)	-	3500	
К	(K)	-	3600	
Л	(L)	-	3700	L' 3701 >RU: сколько;
М	(M)	-	3800	
Н	(N)	-	3900	Ǿ 3901 >CZ: meduňka
П	(P)	-	4000	
Р	(R)	-	4100	
С	(S)	-	4200	
Т	(T)	-	4300	Ǿ- 4301
Ф	(F, PH)	-	4400	
Х	(H)	-	4500	
Ц	(C)	-	4600	
Ч	(CH, Ć)	-	4700	
Ш	(SH, Š)	-	4800	
Щ	(SHT)	-	4900	
	(Ř)	-	5000	>CZ: Například;
	(H)	-	5100	>EN: his, CZ: hmota;
	(DZ)	-	5200	
	(DŽ)	-	5300	
	(Θ)	-	5400	>EN: thank;
	(Ö)	-	5500	>EN: that;
	(N)	-	5600	>EN: sing;

.....

Други;

Тази система за цифрово кодиране на звуците от човешката реч позволява да се запише с цифри звуковият еквивалент на всяка дума от произволен език. За езици извън индоевропейската група, които евентуално притежават звуци, съществено различаващи се от посочените, системата за запис предлага възможности: за гласни от 21 до 29 (още 8 гласни или дифтонги) и от 56 до 99 (още 43 съгласни звука).

Пример:

EN - **tool** = 4300050113700 (туул)
BG - **работа** = 410001013000040043000100
BG диалект **рабутъ** = 410001013000050043000600
RU - **сколько** = 420036000401370136000400

Системата за цифрово кодиране на фонемите позволява в паметта на компютърно устройство да се записват с едни и същи знаци (цифри) звуковите еквиваленти на думите от произволна човешка реч и от произволен език на хомо сапиенс.

ГРАМАТИЧЕСКА КЛАСИФИКАЦИЯ НА ДУМИТЕ

Код А Части на речта

Десети научен факт:

Броят на частите на речта във всички езици е 10.

Това дава възможност при цифрово кодиране да бъде избрана система, при която всяка от частите на речта да бъде кодирана така, че началната цифра да бъде различна (**код А**):

1. **Substantiva** (Съществителни)
2. **Adjektiva** (Прилагателни)
3. **Pronomia** (Местоимения)
4. **Numeralia** (Числителни)

5. Verba (Глаголи)
6. Adverbia (Наречия)
7. Prepozice (Предлози)
8. Conjunctions (Съюзи)
9. Partikule (Частици)
10. Interjekce (Местоимения)

Единадесети научен факт:

Частите на речта във всички езици биват изменяеми и неизменяеми (морфологична класификация).

код В Изменяеми части на речта

- Съществителни
- Прилагателни
- Числителни
- Местоимения
- Глаголи

ГРАМАТИЧЕСКИ КАТЕГОРИИ ПРИ СЪЩЕСТВИТЕЛНИТЕ ИМЕНА

Род, число, определеност (нечленувана и членувана форма), падеж, вид: с една основа (прости съществителни имена), с две и по-вече основи (сложни съществителни имена), броимост (специфично за Английски език), други (ако има такива).

ПЪРВА поредна цифра от код В РОД

Стойност и значение на кода.

- 0 - отсъства род
- 1 - мъжки род
- 2 - женски род
- 3 - среден род
- 4 - общ род
- 5 - притежават два рода (тел, прах)

ВТОРА поредна цифра от код В ЧИСЛО

Стойност и значение на кода:

- 0 - отсъства
- 1 - единствено число
- 2 - множествено число
- 3 - само в единствено число
- 4 - само в множествено число

ТРЕТА И ЧЕТВЪРТА поредни цифри от код В ОПРЕДЕЛЕНОСТ

Стойност и значение на кода:

- 00 - Неопределеност
- 11 - Определеност мъжки род (BG а, DE der, EN а)
- 12 - Определеност мъжки род (BG ът,)
- 13 - Определеност мъжки род (BG я,)
- 14 - Определеност мъжки род (BG ят)

.....

- 21 - Определеност женски род (DE die, BG та)

.....

- 31 - Определеност среден род (BG то)

-
- 41 - Определеност мн. число, мъжки род (BG те)
 - 42 - Определеност мн. число, женски род (BG та)
 - 43 - Определеност мн. число, среден род

.....

- 50 - Обща определеност (EN the, а, an)

Забележка: Формите за определеност, които в някои езици стоят пред съществителните имена : (EN the, а, an, DE der, die, FR le, la, s, un, une, des, les) и др. не са кодират с кодовете за частите на речта (код А). Те не са части на речта.

ПЕТА И ШЕСТА поредни цифри от код В ПАДЕЖ

Стойност и значение на кода:

- 00 - Без падежна форма
- 01 - Именителен падеж

- 02 - Винителен падеж
- 03 - Дателен падеж
- 04 - Родителен падеж
- 05 - Творителен падеж
- 06 - Предложен падеж
- 07 - Звателен падеж
- 08 -
- 09 -
- 10 -

.....

21 до 99 Други падежни форми

Забележка: Езици, които не разполагат с падежни форми (Български, Английски и др.), използват запис „00”.

СЕДМА поредна цифра от код В ВИД С ЕДНА ОСНОВА (ПРОСТИ СЪЩЕСТВИТЕЛНИ ИМЕНА)

Стойност и значение на кода:

- 0 отсъства
- 1 същ. собствени имена
- 2 същ. нарицателни
- 3 същ. с веществено значение
- 4 същ. с отвлечено значение
- 5 същ. събирателни
- 6 същ. за действия
- 7 -
- 8 -
- 9 - други

ОСМА поредна цифра от код В ВИД С ДВЕ И ПО-ВЕЧЕ ОСНОВИ (СЛОЖНИ СЪЩЕСТВИТЕЛНИ ИМЕНА)

Стойност и значение на кода:

- 0 отсъства
- 1 същ. собствени имена
- 2 същ. Наричателни

- 3 същ. с веществено значение
- 4 същ. с отвлечено значение
- 5 същ. събирателни
- 6 същ. за действия
- 7 -
- 8 -
- 9 - други

ДЕВЕТА И ДЕСЕТА поредни цифри от код В БРОИМОСТ (Характерно за Английски език Countable, Uncountable).

Стойност и значение на кода:

- 00 - отсъства
- 01 - неброими имена на материали (material)
- 02 - неброими отвлечени (abstract)
- 10 - броими отделни предмети (individual)
- 11 - броими, група от еднородни предмети, лица

Пример за цифрово кодиране на граматическите категории на съществително име: Код А + Код В:

EN: NEWSPAPER = 10000000310
THE NEWSPAPER = 10050000310

BG: ИНЖЕНЕР = 11100002011
ИНЖЕНЕРЪТ = 11111002011

ГРАМАТИЧЕСКИ КАТЕГОРИИ ПРИ ПРИЛАГАТЕЛНИТЕ ИМЕНА:

Род, число, определеност (нечленувана и членувана форма), падеж, вид: с една основа (прости прилагателни имена), с две и по-вече основи (сложни прилагателни имена), броимост (специфично за Английски език), други (ако има такива).

ПЪРВА поредна цифра от код В РОД

Стойност и значение на кода.

- 0 - отсъства род
- 1 - мъжки род
- 2 - женски род
- 3 - среден род
- 4 - общ род
- 5 - притежават два рода (тел, прах)

ВТОРА поредна цифра от код В ЧИСЛО

Стойност и значение на кода:

- 0 - отсъства
- 1 - единствено число
- 2 - множествено число
- 3 - само в единствено число
- 4 - само в множествено число

ТРЕТА И ЧЕТВЪРТА поредни цифри от код В ОПРЕДЕЛЕНОСТ

Стойност и значение на кода:

- 00 - Неопределеност
- 11 - Определеност мъжки род (BG а, DE der, EN а)
- 12 - Определеност мъжки род (BG ът,)
- 13 - Определеност мъжки род (BG я,)
- 14 - Определеност мъжки род (BG ят)

.....
21 - Определеност женски род (DE die, BG та)

.....
31 - Определеност среден род (BG то)

.....
41 - Определеност мн. число, мъжки род (BG те)

42 - Определеност мн. число, женски род (BG та)

43 - Определеност мн. число, среден род

.....
50 - Обща определеност (EN the, а, an)

Забележка: Формите за определеност, които в някои езици стоят пред съществителните имена : (EN the, a, an, DE der, die, FR le, la, s, un, une, des, les) и др. не са кодират с кодовете за частите на речта (код А). Те не са части на речта.

ПЕТА И ШЕСТА поредни цифри от код В ПАДЕЖ

Стойност и значение на кода:

00 - Без падежна форма

01 - Именителен падеж

02 - Винителен падеж

03 - Дателен падеж

04 - Родителен падеж

05 - Творителен падеж

06 - Предложен падеж

07 - Звателен падеж

08 -

09 -

10 -

.....
21 до 99 Други падежни форми

Забележка: Езици, които не разполагат с падежни форми (Български, Английски и др.), използват запис „00”.

СЕДМА поредна цифра от код В ВИД С ЕДНА ОСНОВА (ПРОСТИ СЪЩЕСТВИТЕЛНИ ИМЕНА)

Стойност и значение на кода:

0 отсъства

1 същ. собствени имена

2 същ. нарицателни

3 същ. с веществуно значение

4 същ. с отвлечено значение

5 същ. събирателни

6 същ. за действия

7 -

8 - други

ОСМА поредна цифра от код В ВИД С ДВЕ И ПО-ВЕЧЕ ОСНОВИ (СЛОЖНИ СЪЩЕСТВИТЕЛНИ ИМЕНА)

Стойност и значение на кода:

- 0 отсъства
- 1 същ. собствени имена
- 2 същ. нарицателни
- 3 същ. с вещественно значение
- 4 същ. с отвлечено значение
- 5 същ. събирателни
- 6 същ. за действия
- 7 -
- 8 -
- 9 - други

ДЕВЕТА И ДЕСЕТА поредни цифри от код В БРОИМОСТ (Характерно за Английски език Countable, Uncountable).

Стойност и значение на кода:

- 00 - отсъства
- 01 - неброими имена на материали (material)
- 02 - неброими отвлечени (abstract)
- 10 - броими отделни предмети (individual)
- 11 - броими, група от еднородни предмети, лица

Пример за цифрово кодиране на граматическите категории на съществително име: Код А + Код В:

EN: NEWSPAPER = 10000000310
THE NEWSPAPER = 10050000310

BG: ИНЖЕНЕР = 11100002011
ИНЖЕНЕРЪТ = 11111002011

ГРАМАТИЧЕСКИ КАТЕГОРИИ ПРИ ПРИЛАГАТЕЛНИТЕ ИМЕНА:

Род, число, определеност (нечленувана и членувана форма), падеж, вид: с една основа (прости прилагателни имена), с две и по-вече основи (сложни прилагателни имена), броимост (специфично за Английски език), други (ако има такива).

ПЪРВА поредна цифра от код В РОД

Стойност и значение на кода.

- 0 отсъства
- 1 мъжки род
- 2 женски род
- 3 среден род

ВТОРА поредна цифра от код В ЧИСЛО

Стойност и значение на кода.

- 0 - отсъства
- 1 - единствено число
- 2 - ,можествено число

ТРЕТА И ЧЕТВЪРТА поредни цифри от код В ОПРЕДЕЛЕНОСТ

Стойност и значение на кода

- 00 - отсъства
- 11 - определеност м. род (BG я)
- 12 - определеност м. род (BG ят)
- 13 - определеност м. род (BG й) (безокий)
- 14 -

.....
21 - определеност ж. род (BG та)

22 -

.....
31 - определеност среден род (BG то)

32 -

- 41 - определеност мн. число мъжки род
 - 42 - Определеност мн. число, женски род
 - 43 - Определеност мн. число, среден род
-

50 - Обща определеност

ПЕТА И ШЕСТА поредни цифри от код В ПАДЕЖ

Стойност и значение на кода:

- 00 - Без падежна форма
 - 01 - Именителен падеж
 - 02 - Винителен падеж
 - 03 - Дателен падеж
 - 04 - Родителен падеж
 - 05 - Творителен падеж
 - 06 - Предложен падеж
 - 07 - Звателен падеж
 - 08 -
 - 09 -
 - 10 -
-

21 - 99 Други падежни форми

Забележка: Езици, които не разполагат с падежни форми (Български, Английски и др.), използват запис „00”.

СЕДМА поредна цифра от код В ВИД С ЕДНА ОСНОВА (ПРОСТИ ПРИЛАГАТЕЛНИ ИМЕНА)

Стойност и значение на кода:

- 0 отсъства
- 1 качествени прилагателни имена
- 2 относителни прилагателни имена
- 3 прилаг. имена от причастия

ОСМА поредна цифра от код В ВИД С ДВЕ И ПО-ВЕЧЕ ОСНОВИ (СЛОЖНИ ПРИЛАГАТЕЛНИ ИМЕНА)

Стойност и значение на кода:

- 0 отсъства
- 1 качествени прилагателни имена
- 2 относителни прилагателни имена
- 3 прилаг. имена от причастия

ДЕВЕТА И ДЕСЕТА поредни цифри от код В БРОИМОСТ (Характерно за Английски език Countable, Uncountable).

Стойност и значение на кода:

- 00 - отсъства
- 01 - неброими имена на материали (material)
- 02 - неброими отвлечени (abstract)
- 10 - броими отделни предмети (individual)
- 11 - броими, група от еднородни предмети, лица

ЕДИНАДЕСЕТА и ДВАНАДЕСЕТА поредни цифри от код В СТЕПЕНУВАНЕ на прилагателните имена.

Стойност и значение на кода:

- 10 - положителна степен
- 20 - сравнителна степен
- 30 - превъзходна степен

Забележка: Думи, които самостоятелно изписани могат да бъдат и съществителни и прилагателни имена, получават цифрови кодове за двете категории части на речта.

Пример за кодиране на прилагателни имена Код А + Код В:
EN: RESPONSIBLE (отговорен) = 2000000100210

ГРАМАТИЧЕСКИ КАТЕГОРИИ ПРИ МЕСТОИМЕНИЯТА

Местоимения съществителни, местоимения прилагателни, местоимения числителни, местоимения наречия.

Лични, притежателни, възвратни и емфатични, показателни, въпросителни, относителни, неопределителни, отрицателни, обобщителни.

Първа поредна цифра от Код В

Стойност и значение на кода:

- 1 - местоимения съществителни
- 2 - местоимения прилагателни
- 3 - местоимения числителни
- 4 - местоимения наречия

Втора поредна цифра от код В -ВИД

Стойност и значение на кода:

- 1 - лични
- 2 - притежателни
- 3 - възвратни и емфатични
- 4 - показателни
- 5 - въпросителни
- 6 - относителни
- 7 - неопределителни
- 8 - отрицателни
- 9 - обобщителни

Трета поредна цифра от код В ЛИЦЕ

Стойност и значение на кода:

- 1 - първо лице
- 2 - второ лице
- 3 - трето лице

Четвърта поредна цифра от код В ЧИСЛО

Стойност и значение на кода:

- 1 - единствено число
- 2 - множествено число

Пета поредна цифра от код В РОД

Стойност и значение на кода:

- 0 - без род
- 1 - мъжки род
- 2 - женски род
- 3 - среден род

Шеста и седма поредна цифра от код В ПАДЕЖ

Стойност и значение на кода:

- 00 - няма падеж
- 01 - именителен падеж
- 02 - винителен
- 03 - дателен
- 04 -
- 05 -

.....
И т.н.

Осма поредна цифра от код В ФОРМА

- 1 - лични местоимения подлози
- 2 - лични местоимения преки допълнения
- 3 - лични местоимения непреки допълнения

Девета поредна цифра от код В ФОРМА

- 0 - пълна форма
- 1 - кратка форма

И т.н.

Пример за кодиране на местоимение:

На_него: 113130030

ГРАМАТИЧЕСКИ КАТЕГОРИИ ПРИ ЧИСЛИТЕЛНИ ИМЕНА:

Първа поредна цифра - ВИД

Величина и значение на кода:

0 - бройни

1 - редни (количествени прилагателни)

Втора поредна цифра - РОД

Величина и значение на кода:

0 - няма род

1 - мъжки род (мъжколична форма)

2 - женски род

3 - среден род

Трета поредна цифра - ЧИСЛО

Величина и значение на кода:

0 - единствено число

1 - множествено число

Четвърта поредна цифра ЧЛЕНУВАНЕ

Величина и значение на кода:

0 - нечленувана форма

1 - членувана форма

И т.н.

Пример:

300

триста : 0010

Членувано:

тристата : 0011

ГРАМАТИЧЕСКИ КАТЕГОРИИ ПРИ ГЛАГОЛИ

Лице, число, род, време, вид, залог, наклонение, спрежение

Първа поредна цифра - ЛИЦЕ

Величина и значение на кода:

- 1 - първо лице
- 2 - второ лице
- 3 - трето лице

Втора поредна цифра - ЧИСЛО

Величина и значение на кода:

- 0 - единствено число
- 1 - множествено число

Трета поредна цифра РОД

Величина и значение на кода:

- 0 - без род
- 1 - мъжки род
- 2 - женски род
- 3 - среден род

Четвърта и пета поредна цифра ВРЕМЕ

(Някои езици имат двуцифрено число времена на глаголите)

Величина и значение на кода:

- 01 - сегашно време
- 02 - сегашно продължително
- 10 - минало свършено
- 11 - минало несвършено
- 12 - минало неопределено
- 13 - минало предварително
- 14 -
- 15 -
- 20 - бъдеще
- 21 - бъдеще в миналото
- 22 - бъдеще предварително
- 23 - бъдеще предварително в миналото
- 24 -

Пета поредна цифра ВИД

Величина и значение на кода:

- 0 - отсъства

- 1 - несвършен
- 2 - свършен

Шеста поредна цифра ЗАЛОГ (активен и пасивен)

Величина и значение на кода:

- 0 - отсъства (безлични глаголи)
- 1 - деятелен залог
- 2 - страдателен залог
- 3 - причастно-страдателни форми

Седма поредна цифра НАКЛОНЕНИЕ

Величина и значение на кода:

- 0 отсъства
- 1 изявително
- 2 преизказно
- 3 повелително
- 4 условно

Осма поредна цифра СПРЕЖЕНИЕ

Величина и значение на кода:

- 0 отсъства
- 1 'е' спрежение
- 2 'и' спрежение
- 3 'а/я' спрежение

Неличните глаголни форми причастия, деепричастия и отглаголни съществителни се кодират едновременно и като глаголи и съответно като прилагателно, наречие и съществително.

Незменяеми части на речта:

НАРЕЧИЯ

Първа поредна цифра - ПРОИЗХОД

Величина и значение на кода:

- 0 - непроизводни
- 1 - производни от имена

- 2 - производни от глаголи
- 3 - производни от прилагателни
- 4 - от падежни форми
- 5 - от числителни имена
- 6 - местоименни наречия

Втора поредна цифра ВИД НА МЕСТОИМЕННИТЕ НАРЕЧИЯ

Величина и значение на кода -ВИД

- 0 - показателни
- 1 - въпросителни
- 2 - относителни
- 3 - неопределителни
- 4 - отрицателни
- 5 - обобщителни

Трета поредна цифра ПРЕДНАЗНАЧЕНИЕ

Величина и значение на кода:

- 0 - за място
- 1 - за време
- 3 - за начин

Четвърта поредна цифра ФОРМА

Величина и значение на кода:

- 0 - пълна форма
- 1 - кратка форма

ГРАМАТИЧЕСКИ КАТЕГОРИИ ПРИ ПРЕДЛОЗИ

Първа поредна цифра ПРОИЗХОД

Величина и значение на кода:

- 0 - първични (прости)
- 1 - производни (сложни)

Втора поредна цифра ДВОЙНИ И УДВОЕНИ

Величина и значение на кода:

- 0 - единични
- 1 - двойни предлози (например: изпод)
- 2 - удвоени (например: за_пред)

Трета и четвърта цифри ЗНАЧЕНИЕ

Величина и значение на кода:

01 -

14 - надвишаване, надминаване на предел

15 - място

.....

И т. н.

Пета поредна цифра СВЪРЗВАНЕ

Величина и значение на кода:

0 - глагол със съществително име

1 - име със съществително име

2 - съществително с наречие

3 - прилагателно със съществително име

КАТЕГОРИИ СЪЮЗИ

Първа поредна цифра ВИД (според морфологичния им строеж)

Величина и значение на кода:

0 - прости

1 - сложни

2 - съотносителни

Втора поредна цифра ФУНКЦИЯ

Величина и значение на кода:

0 - определителни

1 - допълнителни

2 - изяснителни и обстоятелствени

Трета поредна цифра СЪЧИНТЕЛНИ И ПОДЧИНТЕЛНИ

Величина и значение на кода:

0 - съчинителни

1 - подчинителни

Четвърта поредна цифра ВИД НА СЪЧИНТЕЛНИТЕ СЪЮЗИ

Величина и значение на кода:

- 0 - без значение
- 1 - съединителни
- 2 - съотносителни (разделителни)
- 3 - противопоставителни

И т.н.

КАТЕГОРИИ ЧАСТИЦИ

Освен категорията частици тук ще бъдат кодирани и буквите (символите) от съответната азбука.

Първа поредна цифра ФУНКЦИЯ

Величина и значение на кода:

- 0 - без значение
- 1 - със служебна роля в изреченията
- 2 - словообразуващи частици
- 3 - формообразуващи частици

КАТЕГОРИЯ МЕЖДУМЕТИЯ, ПРЕДСТАВКИ и НАДСТАВКИ

Първи три поредни цифри НОМЕРАЦИЯ ПО ВИД

Величина и значение на кода:

- 001 - 199: междуметия
- 200 - 299: представки
- 300 - 399: надставки

ДОПЪЛНЕНИЕ:

Съществуват езици, в които самостоятелно изписани думи могат да бъдат определени едновременно в различни части на речта (съществителни, прилагателни, глаголи и пр.). Тези думи получават цифрови кодове за всяка от частите на речта.

СИНТАКСИС

Код С

Определя ролята на думите в конкретно изречение (след синтактичен анализ).

Можем да използваме код с 4 или по-вече цифри.

Структура на код С

Първи три поредни цифри

Величина и значение на кода:

- 100 подлог
 - 200 казуемо
 - 210 просто казуемо
 - 220 съставно казуемо
 - 221 съставно глаголно казуемо
 - 222 съставно имено казуемо
 - 230 сложно съставно казуемо
 - 231 сложно съставно имено казуемо
 - 232 сложно съставно глаголно казуемо
 - 300 допълнение
 - 310 пряко допълнение
 -
 - 400 обстоятелствено пояснение
 - 410 обстоятелствено пояснение за време
 -
 - 500 определение
 -
- и т.н. до изчепване всички части на изречението.

Четвърта и пета поредни цифри

Величина и значение на кода:

- 01 - първи по ред подлог, казуемо, определение и т.н.
- 02 - втори по ред ...
- 03 - трети по ред ...
-

И т.н. до 99

Код D Вид на изречението

Величина и значение на кода:

- 1 - съобщително (разказно) изречение („.”)
- 2 - въпросително изречение („?”)
- 3 - заповедно изречение („!”)

Код E - Вид на изречението съобразно броя на предикативните центрове

Величина и значение на кода:

- 1000 просто изречение
- 2000 сложно изречение
- 2100 сложно съчинено изречение
- 2110 сложно съчинено съединително изречение
- 2120 сложно съчинено противоположно
-
- 2200 сложно съставно изречение
- 2210 сложно съставно с подложно подчинено
- 2220 сложно съставно с предикативно подчинено
-
- и т.н. до изчерпване на всички видове.

Код AC - ОБЛАСТ НА ПОЗНАНИЕТО (Area of cognition)

С този код разделяме познанието за света на области, в които идиомите имат различен смисъл. Всяка дума, която е носител на качеството „идиома” ще получи отделен код за различните области на познанието.

Цифровият код С е временен и се отнася за конкретното изречение от превеждания текст.

Величина и значение на кода:

- 01 - астрономия
- 02 - биология и медицина
-
- 14 - общо познание
-
- 38 - спорт
- И т.н.

ЕЗИКОВ КОД KL (kind of language)

С този код записваме различните езици и значими диалекти в тях.

Величина и значение на кода:

00100 Албански език

00101 Първи значим диалект в албански език

00102 Втори диалект

00103 Трети диалект...

.....

05000 Английски език

05001 Американски английски

05002 Тексаски диалект

.....

10000 Български език

10001 Родопски диалект

10002 Шопски диалект

.....

60000 Словашки език

60001 Захорацки диалект

60002 Шаришки диалект

60003 Спишки диалект

..... и т.н.

Дванадесети научен факт:

Независимо от езиковия си произход, изречението е основна граматически оформена единица на свързаната реч, която се отличава със смислово и интонационно единство, с комуникативна значимост.

Изобретението предлага метод за превод на завършени изречения от текст.

Преди да се пристъпи към превод, изреченията от текста се подлагат на синтактичен анализ. Така се определят кодовете С, D, E.

Тринадесети научен факт:

При извършване на синтактичен анализ във всички езици

са валидни правила за **УПРАВЛЕНИЕ, СЪГЛАСУВАНЕ, ПРИЛАГАНЕ, ПРЕДЛОЖНО СВЪРЗВАНЕ** и пр., съобразно граматиката на съответния език.

При синтактичния анализ на изречението е необходима и достатъчна информацията от кодовете А и В. Тази информация е достатъчна и за определяне на словоредата на превеждания текст.

Синтактичният анализ започва от т.н. **„Предикативен център”** връзката **ПОДЛОГ - СКАЗУЕМО**. За целите на синтактичния анализ програмата ще търси първо коя дума е глагол в изречението (начална цифра 5). След това кое съществително (начална цифра 1) или лично местоимение и т.н. е съгласувано по род и число с глагола и е подлог в това изречение. Тази информация е скрита в кодовете „В”. По тази логика се прави пълен синтактичен анализ и всяка дума от конкретното изречение получава свой временен код.

Цифровите кодове от един език се трансформират в цифрови кодове на втория език. Прави се отново синтактичен анализ въз основа на временните кодове „С”, прилагат се правилата от синтаксиса на втория език (управление, съгласуване и т.н.) и изречението се възпроизвежда със словоредата на втория език в звук или писан текст.

ФРАЗЕОЛОГИЯ

КОДИРАНЕ НА СЕМАНТИЧНО ЕДНОЗНАЧНИ ИЗРАЗИ ОТ ДВЕ ИЛИ ПО-ВЕЧЕ ДУМИ (ФРАЗЕОЛОГИЧНИ ИЗРАЗИ)

Във всички езици изреченията се състоят от семантично еднозначни отделни думи и/или семантично еднозначни, фразеологични изрази (семантичен сбор от две или по-вече думи) фразеологични сраствания, фразеологични единства, фразеологични словосъчетания. Семантичното значение на тези елементи от фразеологията се кодира като сбор от отделните думи с употребата на свързващ знак,
Например „_”.

Фразеологични сраствания: - Словосъчетание, общото значение на което няма никаква връзка със значението на съставлящите го думи.

Хора _от_ кол _и_ въже.

Фразеологични единства: - Словосъчетание, което се употребява като идиомите в своеобразно значение, но връзката между общото значение и значението на лексическите съставки не се губи.

Хитрата _сврака_ с _двата_ крака.

Фразеологични словосъчетания: - Словосъчетания, които нямат значение, различно от онова на лексичните съставки, но в които една от съставките може да бъде заменена с друга, синонимна, без това да изменя общия смисъл.

Давам _израз_ = Изразявам

Междинни категории: - Фразеологични словосъчетания при изрази, които вече са се откъснали от свободните съчетания, но от друга страна нямат още характер на напълно устойчиви изрази.

Страна _и_ нрави.

Глаголни времена с участието на спомагателни глаголи.

EN:

I _ have _ been _ working.

(The Perfect Progressive Present Positive Tense)

I _ have _ not _ been _ working.

(The Perfect Progressive Present Negative Tense).

Цифровият еквивалент на последните два примера ще кореспондира с цифровите кодове на глагола „WORK работя, занимавам се”, които дефинират съответно: първо лице, единствено число, сегашно продължително време и първо лице, единствено число, сегашно продължително време с отрицание.

I _ have _ been _ working. = 5100020000

Всички тези фразеологични изрази могат да бъдат запомнени като семантично еднозначни структури, които ще бъдат превеждани с една или по-вече думи или съответни фразеологични съчетания в другия език. За нуждите на синтактичния анализ обаче софтуерът ще разглежда отделните думи като самостоятелни единици в състава на изречението.

(Ще пренебрегва знака “_”).

Пример:

EN: It's _ raining _ cats _ and _ dogs.

BG: Вали _ като _ из _ ведро.

Цифрово кодиране на мястото на думите в езика

Най-значимото постижение на човечеството за цялото му хилядолетно съществуване са езиците на отделните народи и изумителното им качество да служат за комуникация. Всяка известна, нововъзникнала или приета от друг език дума е богатство, което заслужава да бъде подредено в сейфа на езика. Великият руски учен Менделеев подреди химичните елементи, които съставят природата в специална периодична система, като им даде пореден номер и остави място за неоткритите още химични елементи.

Тъй като за целите на технологията преследваме възможността цялата човешка реч да бъде представена само с цифри, бихме могли да предложим предварителна система, в която всяка дума да получи свой пореден и единствен номер в речника на езика. Това би могло да бъде и предмет на отделен проект. Числата от математиката ни дават неограничени количествени възможности и затова приемаме две основни категории думи:

·Известни вече думи;

Такива, които биха възникнали по-късно;

Освен това при личните имена, географските означения и др. трябва да се има предвид и възможност думи, които не са от съответния език, да бъдат идентифицирани точно при превод от един език на друг.

Важно е също така да се има предвид, че изборът на система за подреждане на думите от речника на всеки език може да ни предостави възможност отделните части на речта да бъдат групирани и съобразно други признаци, например такива, които са свързани с участието на предлозите времева продължителност, принадлежност, дистрибуция, насоченост и т.н., или признаци като - качество, движение, механизъм, средство, план, въпросителна дума, молба, покана и т.н.

Такава система ще позволи по-лесно намиране на кореспондиращи семантично-еднозначни думи и фразеологични словосъчетания от речника на езика, на който се превежда.

Предвид тези и други съображения за нуждите на технологията предварително и условно приемаме следната система за номериране на думите в основен вид първо лице, единствено число, сегашно време за глаголите (избираме система с шест цифрено номериране, но няма проблем тя да бъде и с по-голям брой цифри седем, осем и т.н.):

· Съществителни нарицателни имена	000001- 099999
· Съществителни собствени имена	100000- 399999
· Географски означения	400000- 499999
· Прилагателни имена	500000- 599999
· Глаголи	600000-699999
· Числителни имена	700000-799999
· Наречия	800000- 899999
· Местоимения	900000-924999
· Предлози	925000-925999
· Съюзи	926000-926999
· Частици (включително буквите)	927000-989999
· Междуметия	990000-999999

В тази система за номериране на думите допускаме, че има място за думи, които ще се появят по-късно в езика. При необходимост системата може да бъде разширена с още цифри (вместо предвидените в един клас макс. 99999 прилагателни имена 999999, 9999999 или по-вече).

Цифровите кодове PD са връзката на цифровата писменост с база данни. Известно е, че преобладаващата част от междуезиковите речници показват думите в основен вид (първо лице, единствено число, сегашно време). Еднократният запис на цифровите кодове PD в паметта на системата ще позволи при превод на текст след всяко изписване на думите да се определи техния код PD а оттам след морфологичен и синтактичен анализ и останалите кодове.

Всички правила за правопис от граматиката на съответния език могат да бъдат въведени като алгоритми с помощта на цифровата писменост.

ПРИМЕР:

Глаголът от Английски език WORK (работя) може да получи PD код за основен вид (първо лице, единствено число, сегашно време): WORK=623125.

Но като изменяема част на речта съгласно граматиката, този глагол може да има форми за лице, число, време и пр. И ако съгласно Английската граматика The Present Participle (Причастието като нелична форма на глагола) се образува като се прибави наставката „-ing” към инфинитива, то това граматическо правило може да бъде записано, и системата да разпознае, че при думата WORKING става въпрос за причастие, което влиза в състава на продължително време (Progressive Tense). Алгоритъмът ще съдържа цифровия код на инфинитива на глагола (код А, съответно с начална цифри 5), плюс цифровия код на надставката (код А, съответно с начална цифра 0 и номера на надставката).

Цифров образ на произволно изречение

Код за вид на езика Код за област на познанието Σ ||Код за мястото на думите в езика Код за съответната част на речта
Код за граматически категории Фонетичен код Код за ролята на думата в изречението || Код за вида на изречението>Код за структура на изречението.

KL AC Σ || PD, A, B, PH, C || D E

Внимание: Фонетичните кодове PH присъстват в цифрово-кодовата система с възможни варианти от разговорната реч.

Пример:	Литературна реч	Отива на работа.
	Диалектна форма	Утивъ нь рабутъ.

Променените фонемии или цели думи се записват предварително заедно с правилните цифрови кодове А и В и програмата за фонетичен превод трябва да ги възприеме и възпроизведе в превод както литературно изговорени. Това ще даде възможност за звукова комуникация от вида: диалект-към-литературен език,

Диалект-към-диалект, литературен език-към-диалект и литературен език-към-литературен език.

Подготвителни дейности за съставяне на софтуер за комуникация на майчин език

- Определяне структурата и състава на кодовете PD (мястов речника). Кодирание с кодовете PD на думите в основен вид. Софтуерът трябва да дава възможност при изписване на думите автоматично да се идентифицират кодовете PD.
- Изготвяне на помощни софтуерни продукти за определяне на кодовете A, B, C, D, E. Всички правила за правопис се записват като алгоритми, представени с цифровите кодове.
- Изготвяне на помощни софтуерни продукти за разпознаване на фонемите от предварително записан глас. Избират се ограничен брой думи, чиито фонемни предварително са кодирани с фонетичните кодове PH. Обхватът от думи съдържа всички възможни позиции на фонемите. Софтуерът трябва да дава възможност за бъдеща произволна комбинация от записаните фонемни, съобразно фонетичните образи на думите от други езици. Липсващите фонемни от един език се допълват от друг глас или синтезатор.
- Изготвяне или представяне на парсърите за синтактичен анализ с помощта на цифровите кодове - Намиране на т.н. „предикативен център”, откриване на глагола-сказуемо посредством код $A = 5$, намиране на подлога, който е съгласуван по (род) и число, словоред и т.н.

Последователност от операции за превод текст-към-текст

(Програмата може да бъде задействана след знаците точка, въпросителна, удивителна и т.н.)

Запис на изречението в паметта на компютъра

Връзка с база данни

Определяне на кодовете PD

Синтактичен анализ

Определяне кодовете A, B и временните кодове C

Определяне кодовете D, E

Предвод на семантично еднозначните единици

Морфологичен и синтактичен анализ

Определяне точните кодове A, B и временните кодове C

Определяне кодовете D, E

Писмено възпроизвеждане на изречението на другия език

Последователност от операции за превод звук-към-звук

Звукозапис на изречението в паметта на компютъра

Определяне комбинациите от съставни фонемите за думите

Определяне цифровите кодове PD

Синтактичен анализ

Определяне кодовете A, B и временните кодове C

Определяне кодовете D, E

Предвод на семантично еднозначните единици

Морфологичен и синтактичен анализ

Определяне точните кодове А, В и временните кодове С

Определяне кодовете D, E

Звуково възпроизвеждане на изречението на другия език

Примерен цифров образ на изречение с превод на няколко езика

Уточнение:

- Цифрите са произволно избрани и показват само общ вид на цифровите кодове;
- Всяка дума е записана и със съответните фонетични кодове, което позволява комуникация текст-към-текст, глас-към-глас, глас-към-текст и обратно;
- Словоредът във всеки език е избран предвид наличието на временните кодове С.

BG: Това изобретение ще промени нашите представи за комуникация! KL code = 10000, AC code = 14

KL AC ||PD, A, B, PH, C || D E

Това (мест.)

∑914238 3 1431300 4300040031000101 500

изобретение (същ.)

023456 1 1100000010 03003500040030004100020043000201
390003000200 100

ще промени (гл.)

618763 5 313200100 49000200_40004100040038000200390
00301 200

нашите (мест.)

904321 3 223220020 390001014800030043000200 310

представи (същ.)

087645 1 2200003410 4000410002003300420043000101310
00300 300

за (предл.)

925114 7 0001 35000101 0000

комуникация! (същ.)

054553 1 2100002011 36000400380005003900030036000
101460003001900 500||
2 100

EN: This invention will change our ideas of communication!

KL code = 05000, AC code = 14

KL	AC	Σ PD, A, B, PH, C	D	E
----	----	--------------------	---	---

This (pronoun)

Σ938244 3 1431300 540003003500 500

invention (noun)

038339 1 1100000010 03003900310002013900480006003
900 100

will_change (verb)

638256 5 313200100 050003003700_4700090039005300
200

our (pronoun)

904321 3 223220020 13010600 310

ideas (noun)

087666 1 2200003410 10003300030001004200 300

of (preposition)

925110 7 0001 06003100 0000

Communication! (noon)

0555543 1 2100002011 3600010038003800050039000300360
00900480006003900 500||
2 100

Подобен цифров вид ще имат и преводите на това изречение на други езици:

DE: Diese Erfindung wird unsere Ideen von der Kommunikation ändern!

FR: Cette invention changera nos idées de communication!

PT: Esta invenção modificará as nossas ideias de comunicação!

RU: Это изобретение изменит наши идеи относительно коммуникации!

ES: ¡Esta invención cambiará nuestras ideas de la comunicación!

Dutch: Deze uitvinding zal onze ideeën van mededeling veranderen!

IT: Questa invenzione cambierà le nostre idee della comunicazione!

Предимства

- Технологията позволява след уточняване на количеството граматически категории за частите на речта в езиците по света, да бъде съставена обща матрица за кодовете А и В. В тази обща матрица всяка клетка може да има данни за съответната граматическа категория на конкретния език.
- Фонетичните кодове могат да бъдат уточнени окончателно от специалисти фонетици така, че за всички езици фонемите да имат общ цифров запис фонетичен код РН.
- Всички правила от граматиката могат да бъдат въведени с помощта на цифровата писменост. Например правилата за „потъмняване” на звуците от разговорната реч могат да бъдат въведени в програмата така, че фонетичният образ на речта да бъде възприеман еднозначно.
- Използването на обща за всички езици цифрова писменост за звуковата и писаната реч, която се свързва с база данни (междуетикови речници) ще позволи върху електронен носител (чип, РС процесор и т.н.) да бъде записана информацията за лексикалния състав на множество езици.
- При отсъствие на достатъчно добри междуетикови речници (например ако няма все още Българо-Малтийски речник в електронен вид), преводът в звук и текст може да бъде осъществен и през трети език например Английски език.

ДРУГИТЕ ЗА НАС

1. ИНСТИТУТ „ДА ВИНЧИ” - USA

Учени изобретиха универсален преводач

<http://www.impactlab.com/modules.php?name=News&file=article&id=2475>

2. Global Watchtower™ - Massachusetts, USA

Българин изобрети уникален метод за превод. Изобретението може да се окаже „Свещения граал” на комуникациите

http://www.commonseadvisory.com/news/global_watchtower_one.php?wat_id=164

3. Agency “FRANCE PRESS”

Des chercheurs bulgares ont mis au point un algorithme de codage numérique de la parole humaine.

http://www.bethel-fr.com/afficher_info.php?id=4656.124

4. VIETNAM PRESS

http://www.daichung.com/110/12_tinhho.shtm

5. GARDIANUL ROMANIA

S-a inventat aparatul care traduce instantaneu

<http://www.gardianul.ro/index.php?a=liniafierbinte2003100702.xml>

6. ITNews Slovakia

Ще промени ли едно гениално българско изобретение комуникацията между хората?

http://www.itnews.sk/buxus_dev/generate_page.php?page_id=37989

7. ИЗВЕСТИЯ Русия

СОЗДАН МГНОВЕННЫЙ СУПЕРПЕРВОДЧИК ДЛЯ ТЕЛЕФОНОВ

<http://www.inauka.ru/news/article36336.html>

И много други...

Invest in this project!
Инвестирайте в този проект!

Благодаря за вниманието!
Thank you for your attention!

February 2006
Copyright © **Koycho Mitev**
E-mail: patentservice@abv.bg