

UOT: 81.322.2

MAYA HEYDƏROVA *

MƏTNİN AVTOMATİK TƏHLİL PROQRAMLARI VƏ ONLARIN İŞ PRİNSİPLƏRİ

XÜLASƏ

Kompüter texnologiyalarının həyatımıza geniş sirayət etməsi dilçilik elminə də öz təsirini göstərmişdir. Bu təsir həm ənənəvi dil problemlərinin öyrənilməsində yeni metod və yanaşmaların ortaya çıxması, həm də dilçiliyin maraq dairəsinin genişlənməsində müşahidə oluna bilər.

Müasir kompüter texnologiyaları böyük həcmli mətnlərin linqvistik təhlili və emalı prosedurlarını xeyli asanlaşdırır və sürətləndirir. Müəyyən tədqiqat prosedurlarının həyata keçirilməsində sərf olunan vaxtın azaldılması daha əhəmiyyətli nəticələrin əldə olunmasına gətirib çıxarır, çünki əvvəllər illər çəkmə məsələlərin həlli bu gün bir neçə saat ərzində mümkündür. Bundan əlavə, İT texnologiyalar linqvistik biliklərin obyektivləşdirilməsinə, onun doğruluğuna kömək edir.

Hal-hazırda dilçilik elminin informasiya texnologiyaları ilə birgə qarşılıqlı əlaqəsi nəticəsində hazırlanmış sadə orfoqrafik səhvləri yoxlama proqramlarından tutmuş axtarış sistemlərində istifadə olunan daha inkişaf etmiş alqoritmlərə, avtomatik referatvermə proqramlarına, maşın tərcüməsinə, ekspert sistemlərinə qədər bir çox tətbiqləri vardır. Məqalədə mətnin avtomatik emalını və təhlilini həyata keçirən kompüter proqramları və onların iş prinsipləri araşdırılmışdır. Dilçiliyə aid müxtəlif məsələlərin həllinin avtomatlaşdırılması, mövcud linqvistik proqramların müxtəlifliyi əks etdirilmişdir.

Açar sözlər: linqvistik təhlil, proqram təminatı, sintaktik analizator, morfoloji analizator, parser,

Giriş. Hazırda insan fəaliyyətinin istənilən sferasında, o cümlədən dilçilikdə İT-nin istifadəsi fəaliyyətin səmərəliliyini əhəmiyyətli dərəcədə artırır. İlk baxışdan İT sahəsi ilə dilçilik sahəsi bir-birinə nə qədər uzaq görünsələr də, aralarındakı danılmaz əlaqə kompüter dilçiliyinin mövcudluğu ilə təsdiqlənir. Dilçiliyin əsas tədqiqat mövzusu mətnlər olduğundan, mətnlərin emalı və təhlili üçün yeni imkanlar açan və mətn məlumatlarının yaradılması, yayılması və axtarışı üçün müxtəlif vasitələri təmin edən müasir İT olduğunu düşünsək, bu iki sahə arasında qarşılıqlı əlaqə olduğunu aydın görürük.

I. Qeyd olunmalıdır ki, İT texnologiyaları dilçiliyin tətbiqi sahələrində daha geniş istifadə olunur ki, kompüter dilçiliyi sahəsində də tətbiqi məsələləri yerinə yetirən xüsusi sinif proqramlar - linqvistik parserlər vardır. Linqvistik parser – mətnin leksemlərini xətti ardıcılıqla təhlilini təmin edən proqram modulları kompleksidir. Eyni zamanda təbii dillərdə cümlənin morfoloji, sintaktik, semantik (bəzən də pragmatik) səviyyədə təhlilini həyata keçirən çoxsəviyyəli təhlildən istifadə olunur.

Bu cür tətbiqi məsələlərə süni dildə (SQL, SPRQL və s.) sorğunun formalaşdırılması və tanınması üçün təbii dildə cümlənin analiz və

sintezi, cümləni bir təbii dildən digərinə çevirmək (maşın tərcüməsi sistemləri), mətnlərin semantik təhlili məsələləri – Text Mining, mətn resurslarının semantik təsnifatı və s. daxildir.

Mətnlərin kompüterdə avtomatik emalı, sintaqmaların morfoloji və sintaktik təhlilinə əsaslanır, cümlələrin formal qrammatika qaydalarına uyğun aparılır. Proqramların işi mətn korpuslarının bazasına əsaslanır.

Kompüter dilçiliyində mətn təhlili bir neçə səviyyədə aparılır:

1. Morfoloji – bir sözün qrammatik əsası vurğulanaraq hansı nitq hissəsinə aid olduğu müəyyənləşdirilir və söz lüğəvi formasına salınır.

2. Qrafematik – mətnin verilənlər massivindən tokenlər seçilir.

3. Sintaktik – cümlədəki sözlər arasındakı sintaktik əlaqələr müəyyənləşdirilir, cümlənin sintaktik quruluşu təyin edilir.

4. Semantik - sözlər və onların məna münasibətləri arasında semantik əlaqələr müəyyən edilir [Ильвовский Д. :<https://www.osp.ru/os/2014/01/13039687>].

Mətn təhlili ilə bağlı linqvistik məsələlərin həlli üçün dilçilik sahəsi üzrə hazırlanmış xüsusi kompüter proqramları vasitəsilə böyük həcmli təkrarlanan əməliyyatlar tələb olunur. Bu proqramlar vasitəsilə nəinki nisbətən sadə (tezliyin təhlil edilməsi, morfoloji və sintaktik

* AMEA Nəsimi adına Dilçilik İnstitutu, Kompüter Dilçiliyi şöbəsi

təhlil), həmçinin nisbətən çətin (semantik təhlil, mətnin üslubunun avtomatik təyini və hətta mümkün müəllifinin təyini) tapşırıqlar yerinə yetirilir. S. Loqiçevin kataloqunda bir çox linqvistik proqram məhsulları təqdim edilmişdir, lakin burada bəzi proqramlar artıq köhnəlmişdir. Şərti olaraq bu proqramlar dörd qrupa bölünür [<http://rvb.ru/soft/catalogue/catalogue.html>]:

1. Lüğətlər və tezauruslar. Kağız formalı analoqlarından fərqli olaraq avtomatlaşdırılmış elektron formalı lüğət proqramlarında axtarış aparmaq daha əlverişlidir. Eyni zamanda xüsusi axtarış vasitələri ilə təchiz edilmiş bu cür resursların müxtəlif variantlarının yaradılması məsələsi aktual olaraq qalır.

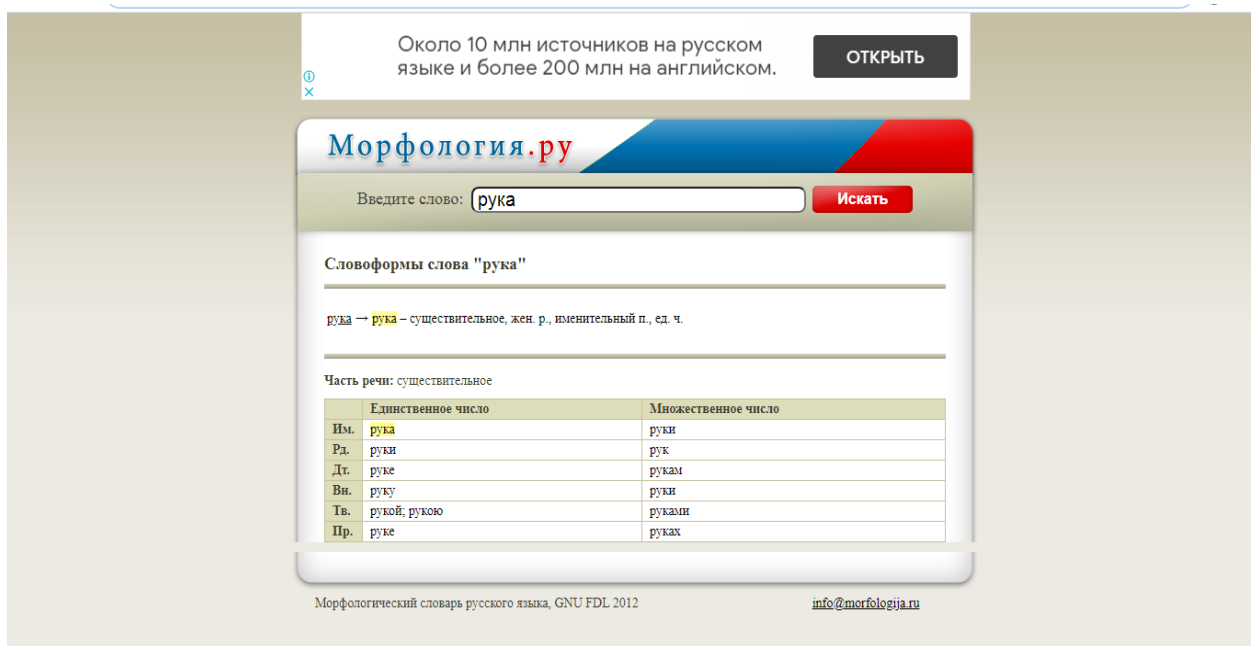
2. Mətn redaktə edən proqramlar və mətn generatorları. Bu sadə tətbiqlər hər hansı mətn redaktorunda quraşdırılır və mətn hissələrini avtomatik redaktəsini (Spell Check) həyata keçirməyə imkan yaradır. Bu proqramlar qrupuna Microsoft Office proqramlar paketinin istifadəçilərinə tanış olan mətnin avtomatik referatlaşdırılması və annotasiyalaşdırılması üçün proqramlar da daxildir. Bununla belə ixtisaslaşdırılmış daha güclü imkanlara malik olan tətbiqlər də vardır. Hal-hazırda veb saytların yaradılmasında və axtarış motorlarının optimallaşdırılmasında geniş şəkildə istifadə olunur. Bu proqramlar müəyyən bir mətnə əsaslanaraq, məndəki sözlərin sinonimlərindən istifadə edərək, onun bir çox təkrarlanan dublikatlarını yaratmağa imkan yaradır. Bu cür proqramlar əvvəlcədən təyin edilmiş şablonlar və ya sinonimlər lüğəti əsasında işləyir. Nümunə kimi SeoGenerator.ru, Generating the Web, SEO Anchor Generator proqramlarını göstərmək olar. Bununla yanaşı, müəyyən bir mövzuda tamamilə müstəqil mətn yaratmaq üçün proqramlar da var. Məsələn, <https://yandex.ru/referats> layihəsi demək olar ki, hər mövzuda mətn yaratmağa imkan verir.

3. Mətnin təhlili və linqvistik emalı proqramları. Bu proqramların sadə versiyaları ofis paketinin istifadəçilərinə məlumdur. Onlar orfoqrafik səhvlərin, mətnin qrammatik və üslubi səhvlərinin yoxlanılmasına, sadə statistik təhlilin aparılmasına imkan yaradır. Lakin bu proqram daim yenilənməlidir. Çünki orfoqrafiya lüğətləri dəyişir, yeni sözlər əlavə olunur, bəzi sözlərin işə yazılışı dəyişdirilir. Proqramda bəzi qrammatik formaların da anlaşılması özünü

göstərir. Bundan başqa konkordansların, tezlik təhlili üçün proqramların hazırlanması tələb olunur. Bu zaman kompüterləşmənin üzərinə daha böyük vəzifələr düşür: sözlərin lemmatizasiya (normallaşdırılması, ilkin formaya gətirilməsi), sintaktik və morfoloji təhlil. Bu tətbiqlərin əsasında bir qayda olaraq stemming alqoritmləri durur. Bundan əlavə, linqvistik proqramların bu qrupuna mətnlərin identifikasiyası və təsnifatı üçün utilitlər tələb olunur. Bunlar da plagiatla bağlı olan məsələləri həll etməyə kömək edir (məsələn, antiplagiat.ru), tanımadığı bir mətnin mümkün müəllifini müəyyənləşdirmək (məsələn, Linqvoanalizator - www.rusf.ru/books/analysis), funksional mətn üslubunun yaradılması (məsələn, teneta.rinet.ru/2001/hudlomer).

4. Təbii dillərin emalı sistemi. Bu qrupa süni intellekt sistemlərinə əsaslanmış nisbətən çətin linqvistik proqram təminatı daxildir. Burada kompüterlər üçün təbii dil interfeysi yaradan proqramlar əsas rol oynayır. Bu tətbiqlərə misal olaraq Siri, Cortana, OKGoogle kimi proqramları göstərmək olar. Eyni zamanda, bu istiqamət neyron şəbəkə texnologiyasının digər tətbiq sahələri kimi indi də fəal şəkildə inkişaf etdirilir.

II. Mətnin avtomatik emalı zamanı istifadə olunan kompüter proqramlarının bir neçəsini nəzərdən keçirək: Rusdilli mətnlərin sintaktik və morfoloji təhlili üçün istifadə edilən “Russian Morphological Dictionary” (<http://faqproject.ru/morphological-dictionary-russiandownload.html>, müəllif – Sergey Sikroskiy) ASCII – kodlaşdırma sistemi ilə yazılmış mətnlərdə işləyir. SWI-Prolog proqramlaşdırma dilində Windows əməliyyat sistemləri üçün hazırlanmışdır və www.ruscorpora.ru saytına əsaslanır [www.ruscorpora.ru]. Bu proqramda özündə 120 000 sözü ehtiva edən A. Zalinzyakın “Грамматический словарь русского языка” (Rus dilinin qrammatik lüğəti) lüğətindən istifadə olunur. Proqram sözlərin qrammatik xüsusiyyətlərini təyin etməyə imkan verir. Məsələn, rus dilində sözlərin qrammatik xüsusiyyətlərini vermək üçün bu lüğət əsasında hazırlanmış <https://www.morfologija.ru> saytında rus dilindəki “pyka” (əl) sözünün morfoloji xüsusiyyətləri belə izah olunur [<https://www.morfologija.ru>]:



Şəkiləndən də gördüyümüz kimi, lüğətdə sözün bütün morfoloji variantları verilir. Lakin A. Zalinznyakin lüğətində bəzi sözlərlə bağlı məhdudiyətlər vardır. Bunlar son dövrün neologizmləri, xüsusi adlar, müqayisəli formalar (məsələn, “постарше”), zərflər (məsələn, “подетски”), defislə yazılan mürəkkəb sözlər və s. Buna görə ədəbi dil sisteminə yeni gələn sözlərin qrammatik mənsubiyyətini proqram vasitəsi ilə müəyyənləşdirməkdə çətinliklər yaranır [К.И. Якубовский və b.: <https://cyberleninka.ru/article/n/obzor-sovremennyh-lingvisticheskikh-tehnologiy-i-sistem>].

“Mystem” A. Zalinznyakin lüğətinə əsasən hazırlanan rusdilli mətnlərin ödənişsiz, sürətli, kompakt morfoloji parseridir. Bu proqramın Windows və Linux əməliyyat sistemləri üçün versiyaları mövcuddur. Mystem proqramı ədəbi normativ rusdilli mətnlərin morfoloji təhlilini həyata keçirir. Lüğətdə olmayan sözlər üçün fərziyyələr şəkilçilərin tezliyinə əsasən yaranır [Якубовский və b.: <https://cyberleninka.ru/article/n/obzor-sovremennyh-lingvisticheskikh-tehnologiy-i-sistem>].

“Рабочее Место Лингвиста” rus dilindən ingilis dilinə (və əksinə) avtomatik tərcümə qurmaq üçün mətn təhlili təklif edir. Bura bir sıra komponentlər daxildir:

1. Rusdilli mətnlərin sintaktik analizatoru;
2. Rus və ingilis dillərindəki mətnlərin morfoloji analizatoru;

3. Müəyyən bir mətn dəsti üçün konkordansların hazırlanması.

Bu proqram C++ proqramlaşdırma dilində hazırlanmışdır və Windows 9x/2000/NT sistemlərində işləyir.

“Morfoloji analizator” rus və ingilis dillərindəki sözlərin morfoloji təhlili üçün proqramın İnternet versiyasıdır. V.Müllərin və A.Zalinznyakin lüğətlərinə əsaslanaraq giriş sözü üçün əsas forma və morfoloji məlumat əldə etməyə imkan verir. “Morfoloji analizatora” istənilən qrammatik formada rus və ingilis dillərində söz daxil etmək olar. Proqramın rus dili üçün aşağıdakı imkanları vardır:

1. İlkin söz formalar (A. Zalinznyakin lüğətinə əsasən);
2. Lüğət məlumatı;
3. Tərcümə;
4. Daxil edilmiş rus dilindəki sözün morfoloji xüsusiyyətləri [<http://www.keva.ru/ling/rus/help.htm>].

Formanın bir neçə mənası olduğu halda təhlil bütün variantlarda aparılır. Proqrama daxil edilmiş formanın təhlili variantlarını əldə etmək imkanı əsas götürülür, çünki bu variantlar morfun dil sisteminə obyektiv yerini müəyyənləşdirməyə zəmin yaradır.

Lingsoft - təbii dillərin emalı üçün proqram təminatıdır (ingilis, alman və bir çox digər dillər üçün).

- Cümlənin qrammatik təhlili;
- Orfoqrafik və qrammatik yoxlamalar;

- Mətnin intellektual qurulması (avtomatik redaktə);
- Mətn məzmununun ümumiləşdirilməsi;
- Suallara cavab yaratmaq;
- Məntiqi nəticə (məndən gizli məna və biliklərin çıxarılması);
- Mətnin səsli tərcüməsi (tərcümə olunmuş mətnin avtomatik səsləndirilməsi);
- Cümlənin video tərcüməsi (tərcümə olunmuş cümlələrin mənasına uyğun video kliplərin göstərilməsi) [<http://www.aot.ru/technology.html>].

AOT (автоматическая обработка текста)-avtomatik mətn emalı sahəsinin mütəxəssisləri tərəfindən hazırlanan layihədir. Bu layihəyə daxildir:

- Mətnin qrafematik təhlilinin modulu;
- Rus, alman və ingilis dilləri üçün morfoloji təhlil komponentləri;
- Omonimliyi avtomatik ləğv etmək üçün modul;
- Mətnin semantik təhlili modulu;
- Linqvistik axtarış sistemi (konkordans);

Müxtəlif tezaurus və sözlüklər [<http://www.aot.ru/technology.html>, <https://rvb.ru/soft/catalogue/c01.html>].

Bu layihə FRAP (Ümumittifaq tərcümə mərkəzinin maşın tərcüməsi laboratoriyası kollektivinin M. Toreza adına MDPXDİ-nin maşın tərcüməsi 1978-1986-cı illər laboratoriyasının kollektivi ilə birgə hazırladığı rusca – fransızca avtomatik tərcümə sistemi) sistemindən götürülmüş təbii dillərin çoxsəviyyəli təqdiminə əsaslanır.

Dil modelini təşkil edən komponentlər, daxil edilmiş mətni bir-birinin ardınca emal edən linqvistik prosessorlardır. Bir prosessorun çıxışı digərinin girişidir. Bu prosessorlar aşağıdakı komponentlərə bölünür:

- Qrafematik təhlil;
- Morfoloji təhlil;
- Sintaktik təhlil;
- Semantik təhlil.

Hər səviyyə üçün özünün təqdimmə dili işləyir. Təqdimmə dili sabitlər və onların birləşməsi qaydalarından ibarətdir. Qrafematik səviyyədə sabitlər qrafematik deskriptorlardır (təsviredicilər) (ЛЕ – лексема (leksem), ЦК – цифровой комплекс (rəqəmsal kompleks) və s.). Morfoloji səviyyədə qrammemlərdir (рд –

родительный падеж (yiyəlik hal), мн – множественное число (cəm). Sintaktik səviyyədə əlaqələr və qrup adlarıdır (ПОДЛ – mübtədə ilə xəbər arasında əlaqə, ПГ – bərelik qrupları). Semantik səviyyədə – semantik kateqoriyalar və münasibətlərdir [<http://www.aot.ru/technology.html>].

Word Tabulator – MS Windows üçün mətn təhlili proqramıdır. Bu proqram vasitəsi ilə müəyyən mətnlər də elementlər indeksini formalaşdırmaq olur. Bu elementlər sözförmələri ola bilər, N – qrammalar (müəyyən ölçülü söz birləşmələri) və ya ifadələr (sintaqmalar). Proqram ilk öncə rus dili üçün hazırlanmışdır, lakin başqa dillər (məsələn, Ukrayna, İslandiya, İsveç) üçün də istifadə oluna bilər. Bu proqram mətnləri həm bir baytlı ASCII kodlaşdırma sistemində, həm də çoxbaytlı UTF-8 sistemində işləyir [<https://rvb.ru/soft/index.html#04>].

Proqram V.İ.Dalyanın 1880 – 1882-ci illərdə nəşr olunmuş lüğətinin ikinci nəşrinə daxil edilmiş kiril əlifbasının rus dilinin ləğv edilmiş I, Ё, Ъ, Ы, V hərfləri də nəzərə alınaraq kiril əlifbasında düzgün şəkildə işləyir. Avropa və ya Skandinaviya dilləri üçün proqram diakritik simvolları düzgün idarə edir. UTF-8 formatında qədim Misir və ya Çin simvolları da daxil olmaqla icazə verilən bütün simvollar işləyir [http://wordtabulator.sourceforge.net/wt_ru.htm].

Mətnin təhlili ilə bağlı proqramlardan biri də “TextAnalyst 2.0” proqramıdır. Bu proqram müəyyən olunmuş leksik vahidlərin tezliyinin ümumiləşdirilmiş şəkildə təqdim olunması, mətnlərdə qruplaşması və öyrənilən nitq materiallarında tədqiq olunması üçün istifadə olunur. Proqramın əsas imkanları bunlardır:

1. sorğunun sözləri ilə məndə olan məna əlaqələrini nəzərə alaraq mətn parçalarının məna axtarışı;
2. mətnin əhatə etdiyi mövzular və alt-mövzular ierarxiyasını quraraq mətnin təhlili;
3. mətn məzmununun hiperlinklə semantik şəbəkənin avtomatik formalaşması ilə təhlili - mətnin semantik quruluşunun mövzular və alt-mövzular ierarxiyası şəklində müəyyənləşdirilməsi;
4. sorğu sözlərinin mətn sözləri ilə gizli semantik əlaqələrini nəzərə alaraq semantik axtarış;
5. mətnin avtomatik referatlaşdırılması;

6. informasiyanın klasterləşdirilməsi - mətn materiallarının tematik siniflər üzrə bölünməsinin təhlili;

7. mətnin hipermətnə çevrilməsi ilə avtomatik indeksləşdirilməsi;

8. hipermətn strukturu və informasiyaya assosiativ daxil olma imkanı ilə tam mətnli birləşmə bazasının formalaşdırılması

[<http://www.analyst.ru/index.php?lang=eng&dir=content/products/&id=ta>].

TextAnalyst 2.0” proqramının komponentləri bunlardır:

1. “TextAnalyst Lib”- hipermətn elektorn kitablar hazırlamaq üçün istifadə olunur.

2. “TextAnalyst SDK”- lemmatizasiya və anlayışların tezlik siyahılarını hazırlamaq üçün istifadə olunur.

Mətnin avtomatik təhlili proqramlarından biri də rusdilli mətnlərin linqvistik emalı üçün istifadə olunan Система Пропись 4.0 proqramıdır.

Proqramın əsas xüsusiyyətləri aşağıdakılardır:

- Köçürmələrin tənzimlənməsi;
- Formalarını nəzərə almaqla mətnə sözlərin axtarışı;

- Mətndəki sözlərin əvəzlənməsi;

- Orfoqrafiyanı yoxlanması;

- Sözün antonim və sinonimlərinin siyahısının hazırlanması;

- Mətnin qrammatik yoxlanması;

- Lüğətdən istifadə etməklə sözün izahı

[<https://www.predictiveanalyticstoday.com/top-free-software-for-text-analysis-text-mining-text-analytics/>].

Mətnin müəyyən bir üsluba aid olub-olmadığını və mətnlərin orijinalıq dərəcəsini müəyyənənləşdirmək üçün lazımi məlumatları toplamağa imkan verən proqramlardan olan “Mətn məlumatlarının axtarışı və təhlili üçün texnologiyalar” -layihəsinin əsas imkanları isə bunlardır:

- Mətnləri təsnifatlaşdırılması;

- Referatların hazırlanması;

- Mətnin axtarışı;

- Mətnin təhlili (sintaktik, morfoloji, semantik);

- Böyük mətn massivləri üçün axtarış vasitələri [<https://www.predictiveanalyticsto->

[day.com/top-free-software-for-text-analysis-text-mining-text-analytics/](https://www.predictiveanalyticstoday.com/top-free-software-for-text-analysis-text-mining-text-analytics/)].

Bunlardan başqa mətnin avtomatik təhlili ilə bağlı Text Analytics tətbiqləri vardır ki, bu zaman strukturlaşdırılmamış mətn məlumatlarını mənalı mətn məlumatına çevrilir. Bu tətbiqlər NLP (natural language processing) alqoritmlərindən istifadə edərək böyük həcmli mətnlərdən məna çıxarır. Bu tətbiqlərə misal olaraq, Apachi OpenNLP, Google Cloud Natural Language API, General Architecture for Text Engineering-GATE, Datumbox, KH Coder, Natural Language Toolkit, KNIME Text Processing, Textable, Text Mining, Text Analytics, Pattern, Aika, Distributed Machine Learning Toolkit və s. göstərmək olar [<https://www.predictiveanalyticstoday.com/top-free-software-for-text-analysis-text-mining-text-analytics/>].

Hazırda İT sahəsinin inkişaf etdiyi bir cəmiyyətdə İT bazarının əsas simaları (Google, IBM, Microsoft) tərəfindən bir çox modellər qəbul edilmişdir. Bununla belə rus dili ilə bağlı sektorda ingilis, Çin, ərəb və Avropa dilləri ilə müqayisədə nəzərəçarpan dərəcədə gerilik hiss olunur. Son illərdə çex, bolqar, rumın, polyak dillərində mətnin işlənməsi sahəsində və korpusların tərtibində fəal tədqiqatlar aparılmışdır. Azərbaycana gəlincə isə burada vəziyyət daha pisdür. Hətta qazax, gürcü dilləri ilə müqayisədə də Azərbaycan dili ilə bağlı proqramlar azlıq təşkil edir. Baxmayaraq ki, bu sahədə ölkəmizdə aparılan işlər hələ Sovet dövründə başlamışdır və kifayət qədər ciddi irəliləyişlər əldə olunmuşdur.

Nəticə: Əlbəttə ki, heç bir proqram linqvistik təhlili bu sahənin mütəxəssisi kimi təhlil edə bilməz. Lakin, burada yazdığımız proqramlar tədqiqatçıya tendensiyalar haqqında nəticəyə gəlməyə imkan yaradacaq və bunun üçün o minimum vaxt sərf edəcəkdir. Bundan əlavə, bu proqramlar hipotezləri daha böyük həcmli material üzərində və əldə edilmiş məlumatların obyektivliyini inamla yoxlamağa imkan verir.

Nəticə olaraq qeyd edə bilərik ki, mətnin təhlili ilə bağlı proqramların imkanları gündəndən artsa da hələ də bu proqramlarda ciddi çatışmazlıqlar nəzərə çarpır və digər kompüter təhlili proqramları kimi sonda mütəxəssis rəyinə ehtiyac duyulur.

ƏDƏBİYYAT

1. <http://rvb.ru/soft/catalogue/catalogue.html> (istinad tarixi: 10.01.2020)
2. http://wordtabulator.sourceforge.net/wt_ru.htm (istinad tarixi: 13.01.2020)
3. <http://www.analyst.ru/index.php?lang=eng&dir=content/products/&id=ta> (istinad tarixi: 18.01.2020)
4. <http://www.aot.ru/technology.html> (istinad tarixi: 9.01.2020)
5. <http://www.keva.ru/ling/rus/help.htm> (istinad tarixi: 05.02.2020)
6. <https://rvb.ru/soft/catalogue/c01.html> (istinad tarixi: 9.01.2020)
7. <https://rvb.ru/soft/index.html#04> (istinad tarixi: 15.01.2020)
8. <https://www.morfologija.ru> (istinad ə tarixi: 11.01.2020)
9. <https://www.predictiveanalyticstoday.com/top-free-software-for-text-analysis-text-mining-text-analytics/> (istinad tarixi: 11.01.2020)
10. www.aot.ru (istinad tarixi: 05.02.2020)
11. www.langsoft.ch (istinad tarixi: 13.01.2020)
12. www.ruscorpura.ru (istinad tarixi: 10.01.2020)
13. Ильвовский Д., Черняк Е. Системы автоматической обработки текстов. Открытые системы. – 2014 – № 1. [Elektron resurs]. URL: <https://www.osp.ru/os/2014/01/13039687> (istinad tarixi: 11.01.2020)
14. К.И. Якубовский, К.А. Якубовская. Обзор современных лингвистических технологий и систем [Elektron resurs]. URL: <https://cyberleninka.ru/article/n/obzor-sovremennyh-lingvisticheskix-tehnologiy-i-sistem> (istinad tarixi: 23.06.2021)

Maya HEYDAROVA

AUTOMATIC TEXT ANALYSIS PROGRAMS AND THEIR WORKING PRINCIPLES

SUMMARY

The widespread use of computer technology in our lives has also affected the linguistics sphere. This effect can be observed both in the emergence of new methods and approaches to the study of traditional language problems, as well as in the expansion of the interest in linguistics.

Modern computer technology greatly simplifies and speeds up the procedures of linguistic analysis and processing of large texts. Reducing the time spent on certain research procedures leads to more significant results, because the solution of problems that took years is possible today in a few hours. In addition, IT technologies help to objectify linguistic knowledge and its accuracy.

Today, in linguistics there are many applications, from simple spelling checking programs to more advanced algorithms used in search engines, automated reference programs, machine translation, and expert systems developed as a result of interaction with information technology. The article investigates computer programs that perform automatic processing, the analysis of text and their working principles. The variety of existing linguistic software is described, automation of the solution of various linguistic tasks.

Key words: linguistic analysis, software, syntactic analyzer, morphology analyzer, morphology analyzer, parser

Мая ГЕЙДАРОВА

ПРОГРАММЫ АВТОМАТИЧЕСКОГО АНАЛИЗА ТЕКСТА
И ПРИНЦИПЫ ИХ РАБОТЫ

РЕЗЮМЕ

Широкое использование компьютерных технологий в нашей жизни повлияло и на сферу лингвистики. Этот эффект наблюдается как в появлении новых методов и подходов к изучению традиционных языковых проблем, так и в расширении круга интересов к лингвистике.

Современные компьютерные технологии значительно упрощают и ускоряют процедуры лингвистического анализа и обработки больших текстов. Сокращение времени, затрачиваемого на определенные исследовательские процедуры, приводит к более значительным результатам, потому что решение проблем, на которые уходили годы, сегодня возможно за несколько часов. Кроме того, ИТ-технологии помогают объективировать лингвистические знания и их точность.

Сегодня в лингвистике существует множество приложений, от простых программ проверки орфографии до более сложных алгоритмов, используемых в поисковых системах, автоматизированных справочных программах, машинном переводе и экспертных системах, разработанных в результате взаимодействия с информационными технологиями. В статье исследуются компьютерные программы, выполняющие автоматическую обработку, анализ текста и принципы их работы. Описано многообразие существующего лингвистического программного обеспечения, автоматизация решения различных лингвистических задач.

Ключевые слова: лингвистический анализ, программного обеспечения, синтаксический анализатор, анализатор морфологии, парсер