

Принципиальная схема решения задачи автодополнения пользовательских поисковых запросов на русском языке и ее анализ

Голяк Ю.Д.

Белорусский государственный университет, Минск

В работе исследуется проблема автоматического дополнения поисковых запросов для информационных систем с интерактивным пользовательским естественно-языковым интерфейсом.

Цель статьи — постановка задачи автодополнения поискового запроса пользователя, ее анализ, принципиальная схема решения и обоснование требуемой для этого функциональности лингвистического процессора.

Материал и методы. Материалом для исследования послужили массивы реальных поисковых запросов, представленные в открытом доступе в известных базах данных. Методология включает в себя статистический анализ и формально-лингвистические методы.

Результаты и их обсуждение. Функциональность автоматического дополнения пользовательского запроса является неотъемлемой частью поисковых систем. Известные решения задачи, как правило, опираются на автоматически формируемую историю проведенного поиска. Однако существует ряд приложений, для которых такое решение по объективным причинам неприемлемо, чем обусловлена необходимость дальнейших исследований в этой области. В работе предлагается формальная постановка задачи автодополнения запроса и принципиальная схема ее решения с учетом предполагаемой «истории» поиска на базе текстов предметной области. Приводятся классификация типов пользовательских запросов и их синтаксических структур, а также функциональность базового лингвистического процессора, необходимого для решения целевой задачи, и требования к дополнительному модулю формирования подсказок для предполагаемых запросов.

Заключение. Предложенный подход к решению целевой задачи позволяет избежать традиционной зависимости от истории поиска и учесть специфику поискового пространства, обеспечивая, что очень важно, высокорелевантную реакцию поисковой системы.

Ключевые слова: автодополнение, поисковый запрос, предиктивный ввод, автоматический лингвистический анализ текста, синтаксическая структура запроса, лингвистический процессор.

(Ученые записки. — 2020. — Том 31. — С. 142–147)

Conceptual Scheme of the Solution for the Problem of User Search Query Autocompletion in Russian and its Analysis

Goliak J.D.

Belarusian State University, Minsk

The paper investigates the problem of automatic search query completion for information systems with an interactive natural-language user interface.

The purpose of the work is the formal definition of the problem of user search query automatic completion and its analysis, creation of the principle scheme of the solution and description of the functionality of a linguistic processor required for the target task.

Material and methods. The materials for the study include the lists of real search queries available for open access. The research methods are statistical analysis and formal linguistic methods.

Findings and their discussion. The functionality of automatic user search query completion is an essential part of search engines. Common solutions of the problem often rely on an automatically gained user search history. However, there are applications that cannot use this solution for objective reasons, which necessitates further research in this area. The paper proposes the formal definition of the problem of automatic completion of a query and principle scheme of its solution

considering the pre-assumed search “history” based on the texts of the domain area. The classification of types of user queries and their syntactic structures, as well as the functionality of the basic linguistic processor required to solve the target problem and the requirements for an additional module for query suggestions generation are described.

Conclusion. The proposed approach to target problem solution gives an opportunity to avoid the traditional dependence on the search history and to take into account the specifics of the search space, providing a highly relevant reaction of the search system.

Key words: Query Type-Ahead, search query, predictive query input, automatic linguistic text analysis, syntactic structure of a query, linguistic processor.

(Scientific notes. – 2020. – Vol. 31. – P. 142–147)

Автодополнение (автоматическое завершение, предиктивный ввод) пользовательского запроса, в дальнейшем QTA (Query Type-Ahead), является необходимой функциональностью практически в любых современных информационных системах с интерактивным пользовательским естественно-языковым (ЕЯ) интерфейсом. Большинство современных поисковых систем, реализующих автодополнение запроса, опираются на автоматически неявно фиксируемую в процессе их эксплуатации историю проведенного поиска в виде списков наиболее частотных пользовательских запросов и/или их наиболее информативных фрагментов [1–3]. Как отмечается в [4], такое решение не всегда является оптимальным, а иногда и вовсе невозможным. Например, в ряде сфер история пользовательских запросов может быть конфиденциальной информацией. Поэтому актуальным для рассматриваемой задачи является случай, когда поисковое пространство, во-первых, заранее «известно» поисковой системе и, во-вторых, представляет собой относительно постоянную полнотекстовую базу данных (ПБД), например, БД текстов некоторой предметной области. Как показали проведенные исследования, эффективным решением задачи здесь представляется идея предварительного, т.е. до эксплуатации системы, формирования «истории», но не уже осуществленного, а предполагаемого поиска, в виде множества P так называемых подсказок, автоматически распознаваемых в самой ПБД, поскольку именно ей будут адресованы пользовательские запросы.

Цель статьи – постановка задачи автодополнения поискового запроса пользователя, ее анализ, принципиальная схема решения и обоснование требуемой для этого функциональности лингвистического процессора.

Материал и методы. Материалом для исследования послужили списки поисковых запросов к базе Topic Explorer, представленные в [5], (в количестве 8 тысяч запросов); списки запросов, предложенные в [6] для решения задачи вложенной сегментации запроса (в количестве 10 тысяч запросов), и базы запросов, предоставленные в качестве данных для решения задач из области ранжирования запроса на соревновательной платформе kaggle.com [7] (в количестве 10 тысяч запросов). Указанные источники в оригинале содержат англоязычные запросы, однако мы исходим

из предположения о том, что общее представление об их длине, синтаксической и особенно семантической структуре можно, в определенной степени, распространить и на случай русского языка, являющегося в данном контексте фокусом нашего исследования. Информация для анализа на собственно русскоязычном материале была почерпнута в процессе работы с сервисом wordstat.yandex.ru [8], позволяющем получить статистику запросов в Яндексе, включающих некоторое заранее заданное слово или словосочетание, и похожих на них запросов. Использованные методы – статистический экспертный анализ и методы компьютерной лингвистики.

Результаты и их обсуждение. Пусть вводимый пользователем запрос в определенный момент времени представляет собой цепочку

$$W_1 W_2 \dots W_n, \quad (1)$$

где каждое W_i , $i = 1, \dots, n-1$, $n > 1$, является словом естественного языка, а W_n , $n \geq 1$ – словом либо префиксом слова. Задача автодополнения запроса вида (1), в определенной степени уже отображающего информационную потребность пользователя, заключается в автоматическом формировании списка таких, в идеале, грамматически и семантически корректных цепочек слов естественного языка, которые включают в себя члены цепочки (1), то есть речь идет не только о последовательном завершении строки вводимого запроса, рассматриваемого в качестве его префикса, но о более сложной процедуре, предполагающей, в общем случае, погружение уже набранной пользователем части запроса в контекст, а именно дополнение ее в начале, в конце и даже внутри. При этом предлагаемые цепочки могут содержать не все слова из (1). В любом случае они должны быть отранжированы по их релевантности относительно исходной цепочки. Кроме того, процедура автодополнения может применяться к вводимому запросу неоднократно (фрагментарно).

Исходя из того, что пользователи в большинстве своем следуют уже сложившейся практике формирования поисковых запросов и что они традиционно в той или иной степени ориентированы на ключевой поиск, может быть предложена следующая принципиальная схема решения задачи построения множества

подсказок для автодополнения пользовательских запросов в рассматриваемом случае:

- экспертный анализ на основе общедоступных источников наиболее частотных поисковых запросов и классификация их типов;
- классификация синтаксических структур запросов для каждого из их типов;
- автоматический лингвистический анализ ПБД с целью распознавания в ее текстовых документах полученных на предыдущем этапе синтаксических структур и выбор из ПБД всевозможных соответствующих им лексических наполнений, что и составляет множество Р.

Классификация типов поисковых запросов. Проведенный экспертный анализ показал, что наиболее частотные поисковые запросы можно разделить на следующие основные типы:

- одно или несколько несогласованных ключевых слов (*грассирование; купить телефон Минск; Белавиа билет; отдых январь; Миоры погода; распродажи обуви Минск*);
- короткие грамматически согласованные словосочетания (*извержение вулкана в Исландии; срочная починка телефона; выбрать недорогой подарок; уровень пробок на дорогах; сделать нотариальный перевод на белорусский*);
- вопросительные предложения (*как настроить будильник на телефоне?; где находятся Азоры?; каковы причины резкой поломки двигателя?; что такое контргайка?; как предотвратить коррозию металла?*);
- утвердительные предложения (*ноутбук перегревается и шумит кулер; болит голова после сна; холодильник слабо охлаждает; не приживаются саженцы персика; быстро разряжается батарея телефона*);
- комбинация двух последних типов, часто без знака препинания между ними (*ноутбук перегревается и шумит кулер что делать?; болит голова после сна как лечить?; холодильник слабо охлаждает, какого мастера вызвать?; не приживаются саженцы персика, какой нужен климат?; быстро разряжается батарея телефона: чинить или под замену?*).

Классификация синтаксических структур запросов каждого типа. Анализ запросов каждого типа (кроме 1) позволил дать следующую классификацию их синтаксических структур.

Словосочетания:

- именные группы: *лазерная хирургия, хирургия глаза, изготовление ключей, ареал обитания тушканчиков*;
- именные группы с предложно-падежными зависимыми конструкциями: *добыча руды в шахтах; антикоррупционные законопроекты при парламентской системе; осложнения после лазерной коррекции зрения; просмотр сериалов без регистрации*;

- именные группы с причастным оборотом: *заболевания, передающиеся воздушно-капельным путем; средства, понижающие температуру; картины, проданные за самые высокие цены; акции, выросшие в цене за последнее время*;
- глагольные группы с прямым объектом: *купить телефон, забронировать отель; отремонтировать обувь, скачать реферат*;
- глагольные группы с косвенным объектом: *выжить при низкой температуре, выспаться за два часа, выиграть в казино, подготовиться к ЦТ*;
- глагольные группы с прямым и косвенным объектами: *лечить грипп при помощи народных средств, вывести пятно от вина; сделать прививку от гриппа, вырастить черешню на даче*;
- глагольные группы с причастным оборотом: *встроить деталь, снабженную витой резьбой; восстановить гладкость поверхности, поврежденной механически; купить стиральную машину, оснащенную системой самоочистки; восстановить данные с диска, который был отформатирован*.

Вопросительные предложения:

- простые: *как лечить простуду у ребенка?; где найти японский рис в Минске?; чем можно предотвратить коррозию металла?; куда отправить ребенка летом?*;
- сложносочиненные: *почему появляется ржавчина и темнеет металл?; как варить лапшу и жарить курицу для тайских блюд?; как избавиться от пятен, но сохранить цвет ткани?; сколько стоит новый айфон и где его купить?*;
- сложноподчиненные: *как открыть консервную банку в походе, когда нет ножа?; как отличить поганку от боровика, если не разбираешься в грибах?; как дрессировать собаку, которая попала в дом взрослой?; как завести автомобиль, если кончилось топливо?*.

Утвердительные предложения:

- простые: *перегревается двигатель; поднимается температура у ребенка; перестала плодоносить яблоня; продается квартира*;
- сложносочиненные: *шумит кулер и греется системный блок; аккумулятор работает, но автомобиль не заводится; купил термос, а он не держит температуру; наушники подключены, но звука нет*;
- сложноподчиненные: *скрипит дверь, даже если смазать машинным маслом; мигает индикатор разрядки батареи, когда аккумулятор заряжен; хочу вернуть товар, который был куплен больше месяца назад; не принимают товар по гарантии, хотя гарантийный срок еще не истек*.

Комбинация двух последних типов:

- сочетание двух простых предложений, нередко без знака препинания между ними: *протекает кран вызвать сантехника; разбился телефон найти ремонт; аренда машины в минске где и как оформить; болит голова к какому врачу идти*.

Очевидно, что лексическое наполнение именно таких синтаксических структур в той или иной степени будет выражать информационную потребность пользователя с помощью его поискового запроса. И если это наполнение «подтолкнет» его здесь к использованию лексических фрагментов из ПБД, это обеспечит на этапе поиска гарантированно релевантную реакцию системы.

Ориентируясь на полученную классификацию синтаксических структур запросов основных типов, можно дать их формальное определение в терминах металингвистических переменных с целью последующего построения паттернов для алгоритма автоматического распознавания потенциальных подсказок в предполагаемом поисковом пространстве и построения их списка при условии, что существует или может быть построен лингвистический процессор, оперирующий при автоматическом анализе текстов из указанного пространства теми же металингвистическими переменными.

То есть речь идет о так называемом базовом лингвистическом процессоре (БЛП) [9], который является необходимой составляющей для решения большого числа задач в области автоматической обработки текста [10]. Проведенные исследования показали, что для решения целевой задачи в качестве такового может быть использован доступный нам известный многоязычный БЛП IHS Goldfire (далее просто БЛП), поддерживающий работу с текстами на шести языках: английском, немецком, французском, японском, китайском и русском [11].

Функциональность базового лингвистического процессора. Процедура автоматического лингвистического анализа с помощью БЛП предполагает поэтапную обработку текста, причем на каждом из этих этапов входное представление текста определенным образом обрабатывается и обогащается лингвистической информацией. Описываемая функциональность опирается на лингвистическую базу знаний (ЛБЗ), которая, как правило, включает в себя декларативные и процедурные знания о ЕЯ. Как отмечается в [12], к декларативным знаниям относят словари различного назначения, классификаторы свойств ЕЯ на различных уровнях его глубины, статистические данные о ЕЯ и т.п., а процедурную часть ЛБЗ составляют так называемые лингвистические правила (паттерны) различного рода, которые в совокупности и в сочетании с декларативной частью ЛБЗ представляют собой определенную лингвистическую модель ЕЯ.

Рассмотрим этапы такой обработки на примере следующего предложения, извлеченного из документа в формате PDF:

В большинстве случаев влаго- отталкивающие покрытия помогают предотвратить повреждения металла, вызываемые коррозией (в т.ч. «жёлтую» ржавчину - хим. ф. FeO(OH)H₂O).

Этап 1. Форматирование и нормализация текста. Необходимость первого вспомогательного этапа

лингвистического анализа (ЛА) обусловлена существованием различных форматов документов, и для упрощения процесса их обработки производится преобразование текста в некоторый единый формат, максимально сохраняющий стилистическую и структурную разметку документов. Кроме того, на данном этапе осуществляются разбиение текста на параграфы, выделение заголовков, подзаголовков и разделов, а также производится фильтрация вспомогательного текста (в случае обработки интернет-документов к вспомогательному тексту можно отнести, например, тексты кнопок, меню и т.д.). Этот модуль унифицирует разрозненные символы, несущие одно и то же значение в тексте (например, в нашем предложении он приведет к единому виду кавычки разного начертания, удалит лишние пробелы), исправляет неканоничные написания сокращений (*т.ч.* будет приведено к *т.ч.*), а также восстанавливает слова, поврежденные переносом между строками в исходном тексте (*влаго- отталкивающие* будет восстановлено к *влагоотталкивающие*, *корро- зией* – к *коррозией*). Таким образом, на выходе этапа 1 исходное предложение будет выглядеть следующим образом:

В большинстве случаев влагоотталкивающие покрытия помогают предотвратить повреждения металла, вызываемые коррозией (в т.ч. «жёлтую» ржавчину – хим. ф. FeO(OH)H₂O).

Этап 2. Лексический анализ текста. На этапе лексического анализа текста прежде всего распознаются границы его слов и предложений. Здесь же частично или полностью решаются задачи распознавания имен собственных, аббревиатур, электронных адресов, цифровых и других знаковых комплексов. Этот вид ЛА иначе называют сегментацией текста. В большинстве случаев границы слов определяются правилами, основанными на учете пробелов и знаков препинания, а предложений – знаков препинания и следующего за ним слова, написанного с прописной буквы. Однако на практике далеко не всегда можно опереться на эти признаки. Ярким примером в этом отношении могут послужить в рассматриваемом предложении сокращения с точками и формула, содержащая в себе скобки, которые обычно являются отдельным «словом», но в этом случае должны остаться внутри неразрывной единицы текста, за которой следует уже самостоятельная закрывающая скобка: '*FeO(OH)H₂O*'.

В результате этого этапа обработки в предложении будут определены следующие словарные элементы:

<i>В</i>	<i>(в</i>
<i>большинстве</i>	<i>т.ч.</i>
<i>случаев влагоотталкивающие</i>	<i>«жёлтую»</i>
<i>покрытия</i>	<i>ржавчину</i>
<i>помогают предотвратить</i>	<i>–</i>
<i>повреждения</i>	<i>хим.</i>
<i>металла, вызываемые</i>	<i>ф.</i>
<i>высокой</i>	<i>FeO(OH)H₂O).</i>
<i>влажностью</i>	

Этап 3. Лексико-грамматический анализ текста.

Задачей лексико-грамматического анализа текста является определение лексико-грамматической категории каждого его слова с учетом его морфологии и контекста. Множество всех лексико-грамматических категорий ЕЯ обычно задается заранее разработанным классификатором его лексико-грамматических свойств, основанным на разделении слов на категории, обозначаемые тегами, на основании принадлежности к определенной части речи. Рассматриваемое предложение в тегированном виде будет выглядеть следующим образом:

В_IN большинстве_NCNR случаев_NCPG влавоотталкивающие_JPO покрытия_NCPO помогают_V3P предотвратить_VB повреждения_NCPA металла_NCMG ,_ , вызываемые_LPPA коррозией_NCFI ,_ (_LQ в_IN т.ч._NC «_LQ желтую_JFA »_RQ ржавчину_NCFI -_ : хим._NC ф._NC FeO(OH)H2O_FO)_RQ ._,

где, например, тег *IN* обозначает предлог, *NCNR* – нарицательное существительное среднего рода в предложном падеже, *NCPG* – нарицательное существительное множественного числа в родительном падеже, *VB* – инфинитив глагола и т.д. Полный список лексико-грамматических тегов для русского языка в рассматриваемом БЛП включает 247 тегов.

Этап 4. Синтаксический анализ текста. На этапе синтаксического анализа осуществляется распознавание в каждом предложении синтаксических отношений и представление их, как правило, в виде

функционального или синтаксического дерева, в котором для слов предложения указывается их грамматическая функция и определяется тип синтаксической связи между ними. В этот момент из предложения извлекаются канонизированные именные группы (*большинство_NCPO случаев_NCPG, коррозия_NCFI, повреждения_NCPO металла_NCMG, влавоотталкивающие_JPO покрытия_NCPO*), а также глагольные группы с 14-компонентной структурой, в данном случае единственная такая группа (табл. 1).

Уже сами синтаксические деревья могут быть полезны для решения многих задач в сфере автоматической обработки текста. В случае решения задачи автодополнения запроса такие их компоненты, как, например, простые именные группы, являются, очевидно, основой формирования подсказок типов **единичные ключевые слова и именные словосочетания**, а глагольные группы – практически всех остальных типов подсказок, включающих предикат.

Этап 5. Семантический анализ текста. На уровне семантического анализа БЛП распознает отношения между концептами, выраженными именными группами, в рамках так называемой CAO-структуры: Субъект – Акция (Предикат) – Объект. Причём, каждый элемент в этой структуре может иметь атрибуты [9; 12]. В общем случае структура такого отношения состоит из семи полей (плюс служебное восьмое поле). Любое из полей может оставаться пустым в

Таблица 1

1. Подлежащее 1 (с атрибутом)	влавоотталкивающие_JPO покрытия_NCPO
2. Сказуемое 1	помогают_V3P предотвратить_VB
3. Именная часть сказуемого 1	–
4. Прямое дополнение 1	повреждения_NCPA металла_NCMG
5. Предлог 1	в_IN
6. Косвенное дополнение 1	большинстве_NCNR случаев_NCPG
7. обстоятельство 1	–
8. Подлежащее 2	–
9. Сказуемое 2	вызываемые_LPPA
10. Именная часть сказуемого 2	–
11. Прямое дополнение 2	–
12. Предлог 2	–
13. Косвенное дополнение 2	коррозией_NCFI
14. обстоятельство 2	–

Таблица 2

CAO-структура	CAO 1	CAO 2
1. Субъект	влавоотталкивающие покрытия	коррозия
2. Акция	помогать предотвратить	вызывать
3. Объект	повреждения металла	повреждения металла
4. Атрибут прилагательное	–	–
5. Предлог	в	–
6. Непрямой объект	большинстве случаев	–
7. Атрибут наречие	–	–
8. Оригинальное представление акции в тексте	помогают предотвратить	вызываемые

зависимости от наполнения конкретного предложения. Структура двух семантических отношений САО из рассматриваемого примера выглядит следующим образом: табл. 2.

Важно заметить, что такие структуры действительно отражают уже не столько синтаксис предложения, сколько семантические отношения в нем: выполняя в пассивном залоге синтаксическую роль непрямого объекта, *коррозией_NCFI* по сути является субъектом относительно акции *вызывать_VB*, выраженной в предложении при помощи причастия, что и отражается во второй САО-структуре.

Полученные структуры служат основой для формирования будущих подсказок, поскольку:

- канонизированные **именные группы** в полях **субъект** и **объект** САО-структуры по сути являются уже готовыми подсказками: *коррозия, повреждения металла, влагоотталкивающие покрытия* (именная группа *большинство случаев* должна быть отфильтрована как нерелевантная);
- из **прямого дополнения 1**, **сказуемого 2** и **косвенного дополнения 2** синтаксической структуры можно «собрать» подсказку вида **именная группа с причастным оборотом** *повреждения металла, вызываемые коррозией*;
- из поля **акция** второй САО-структуры, где причастие *вызываемые* приведено к начальной глагольной форме *вызывать*, и **объекта** той же структуры конструируется подсказка вида **глагольная группа с прямым объектом** *вызывать повреждения металла*, которая станет подходящим дополнением к запросу, например, начинающемуся с префикса *‘что может’*.

Подразумеваемые при этом преобразования могут быть описаны в виде соответствующих лингвистических правил (паттернов), реализация которых требует дополнительной функциональности БЛП. Она же потребуется, как показал проведенный анализ, и при формировании следующих видов подсказок:

- **сказуемое 1** и **прямое дополнение 1** синтаксической структуры можно объединить для получения подсказки вида **глагольная группа с прямым объектом** (*помогают предотвратить повреждения металла*), но личная форма глагола *помогают* кажется не уместной для запроса. Здесь требуется приведение глагола к форме единственного числа *‘что’ помогает предотвратить повреждения металла* или возможность оставить в подсказке только смысловой глагол из предиката *предотвратить повреждения металла*;
- вторую САО структуру, за исключением служебного поля, целесообразно использовать в качестве подсказки вида **простое утвердительное предложение** *коррозия вызывает повреждения металла*, однако начальная форма глагола не позволяет сделать эту конструкцию согласованной и требует приведения глагола к нужной форме *вызывает*.

Заключение. Таким образом, формирование подсказок на основе описанных компонентов синтакси-

ческих и семантических структур возможно из любого предложения ПБД при соблюдении определенных условий, таких как достаточная информативность компонентов указанных структур, их лексико-грамматическое соответствие требованиям синтаксических типов подсказок, вхождение в заданные семантические группы, грамматическая согласованность компонентов, отсутствие личных имен и иных персональных данных, исключение обценной и ненормативной лексики, ограничение по длине подсказки и др. Безусловно, в состав дополнительной функциональности БЛП необходимо включить возможность ранжирования и классификации компонентов по определенным грамматическим и семантическим критериям. Все это в совокупности с БЛП и составляет лингвистический процессор решаемой задачи.

Литература

1. Incremental query refinement: US Patent 9245052 B2 / Brett D. Brewer, Oliver Hurst-Hiller (2012/2014).
2. Search entry system with query log autocomplete: US6564213B1 / Eric Paul Burke, Duke Fan, Alan Wada, Jawahar Malhotra, Brian Coe. (2005/2008).
3. Search query suggestions based on personal information: US Patent 9317585 B2 / Maureen Heymans, Ashutosh Shukla, Harish Rajamani, Matthew E. Kulick, Bryan C. Horling, Jennifer E. Fernquist, Weniger". (2013/2014).
4. Голяк, Ю.Д. Автодополнение поискового запроса на основе автоматического извлечения подсказок из проиндексированных документов предметной области / Ю.Д. Голяк // Вести БГПУ. Сер. 1, Педагогика. Психология. Филология. – 2018. – № 3. – С. 91–95.
5. Reidsma, M. (2016). Summon Topic Explorer Results by Search Query [Data set]. Zenodo. <http://doi.org/10.5281/zenodo.47723> [Electronic resource]. – Mode of access: <https://zenodo.org/record/47723#.XS4aS-gzaHs>. – Date of access: 07.04.2020.
6. Rishiraj. (2018). Supplementary Material for Nested Segmentation of Web Search Queries [Data set]. Zenodo. <http://doi.org/10.5281/zenodo.1137746> [Electronic resource]. – Mode of access: <https://zenodo.org/record/1137746#.XS4iNegzaHs>. Date of access: 07.04.2020.
7. Krapamih: search-queries [Electronic resource]. – Mode of access: <https://www.kaggle.com/krapamih/search-queries#searchterms.txt>. – Date of access: 07.04.2020.
8. Wordstat.Yandex [Electronic resource]. – Mode of access: <https://wordstat.yandex.ru/>. – Date of access: 07.04.2020.
9. Чеусов, А.В. Разработка алгоритмов и технологии построения многоязычного базового лингвистического процессора: дис. ... к-та техн. наук: 05.13.17 / А.В. Чеусов. – Минск, 2013. – 116 л.
10. System and method for automatic semantic labeling of natural language texts: Patent EP2406731A4 / J. Todhunter, I. Sovpel, D. Pastanohau. (2010).
11. IHS Goldfire [Electronic resource]. – Mode of access: https://www.ihf.com/pdf/IHS-Goldfire-Platform-Whitepaper_140823110915517432.pdf. – Date of access: 07.04.2020.
12. Чернышевич, М.В. Система автоматического сентимент-анализа текстов на английском языке: дис. ... к-та филол. наук: 10.02.21 / М.В. Чернышевич. – Минск, 2019. – 150 л.

Поступила в редакцию 21.04.2020