

Відомості про авторів / Сведения об авторах / About the Authors

Бабкова Надія Вікторівна – кандидат технічних наук, Національний технічний університет «Харківський політехнічний інститут», старший викладач кафедри інтелектуальних комп'ютерних систем; тел.: (050) 712–77–02; e-mail: nadjenna@gmail.com.

Бабкова Надежда Викторовна – кандидат технических наук, Национальный технический университет «Харьковский политехнический институт», старший преподаватель кафедры интеллектуальных компьютерных систем; тел.: (050)712-77-02; e-mail: nadjenna@gmail.com.

Babkova Nadiia Viktorivna - Candidate of Engineering Sciences, National Technical University "Kharkiv Polytechnic Institute", teacher of the Intelligent Computer Systems Department; tel.: (050) 712–77–02; E-mail: nadjenna@gmail.com.

Колодна Лілія Михайлівна –Національний технічний університет «Харківський політехнічний інститут», студентка; тел.: (063) 360–96–60; e-mail: ikskafedra@ukr.net.

Колодная Лилия Михайловна –Национальный технический университет «Харьковский политехнический институт», студентка; тел.: (063) 360–96–60; e-mail: ikskafedra@ukr.net.

Kolodna Liliia Mykhailivna –National Technical University "Kharkiv Polytechnic Institute", student; tel.: (063) 360–96–60; e-mail: ikskafedra@ukr.net.

УДК 004.91

Н. С. УГОЛЬНИКОВА, М. В. ЧУХНЕНКО**ОСОБЛИВОСТІ КЛАСИФІКАЦІЇ ТВОРІВ МАСОВОЇ ЛІТЕРАТУРИ**

У статті наведено опис жанрів сьогоденної масової літератури, приведено аналіз особливостей кожного з розглянутих жанрів. Показано можливість використання частотного словника як засобу визначення формальних жанрових ознак у творах масової літератури. Особлива увага приділена можливості автоматизувати створення частотних словників для визначення жанрів. Розроблено алгоритм автоматичного вилучення частотних слів, які характеризують відповідний текст, з тесту обраного експертом жанру.

Ключові слова: масова література, частотний словник, детектив, наукова фантастика, фентезі, трилер, пригодницький роман.

В статье проведен обзор жанров современной массовой литературы, сделан анализ особенностей каждого из рассмотренных жанров. Показана возможность использования частотного словаря как средства определения формальных жанровых признаков в произведениях массовой литературы. Особое внимание уделено возможностям автоматизировать создание частотных словарей для определения жанров. Разработан алгоритм автоматического извлечения частотных слов, характеризующих соответствующий текст, из текста выбранного экспертом жанра.

Ключевые слова: массовая литература, частотный словарь, детектив, научная фантастика, фэнтези, триллер, приключенческий роман.

The article explores the approach of dividing the modern literary process into a triad and singles out one of its facets – mass literature. Analyzed the differences in the term "literary genre", an overview of the genres of today's mass literature, an analysis of the characteristics of each of the genres examined. The notion "frequency dictionary" is considered in the context of applied linguistics and problems that arise in the construction of frequency dictionaries of large text boxes are described. The possibility of using a frequency dictionary as a means of determining formal genre features in works of mass literature is shown. Particular attention is paid to the ability to automate the creation of frequency dictionaries for the definition of genres. An algorithm for automatically extracting frequency words characterizing the corresponding text from the text of the genre chosen by the expert is proposed, which will automate the process of determining the genre of an arbitrary text.

Keywords: Mass literature, frequency dictionary, detective, science fiction, fantasy, thriller, adventure novel.

Вступ. Література, яка завжди адресувалася представникам різних станів, приречена була ділитися на «високу» і «низову». «Висока» створювалася для вищих у культурному плані станів – людей освічених, підготовлених до читання, які знають традицію, вимогливих. «Низова» література, відповідно, була розрахована на нижчі (в культурному плані) шари, тобто на невибагливих, малоосвічених читачів. Масова культура у сучасному розумінні остаточно формується на межі XIX–XX століть, коли у результаті складних соціальних (посилення міграції, домінування міської культури, демократизація, тощо), науково-технічних (винахід нових форм комунікації) процесів формується феномен масового суспільства, суспільства мас.

В епоху масовості і «технічної відтворюваності» «низова» література в більшій своїй частині зводиться

до літератури масової. До певної міри можна говорити про те, що у середовищі малоосвічених верств населення масова література в якійсь мірі замінює фольклор і навіть сама стає його джерелом. Поступово «споживачами» «низової» літератури, яка активно розвивається, стають нові верстви населення [1].

До кінця XIX століття в літературному процесі складається тріада: авангардна література – зорієнтована на «класичну» літературу белетристика – масова література. Ця «розмітка» літератури продовжує існувати донині. Втім, крайні частини тріади встигли за цей час обзавестися своєю «класикою», і сьогодні абсолютно коректно вживати вирази «класика авангарду», «класика наукової фантастики» або «класика детективу». Традиційно авангард і белетристику відносять до елітарної

літератури, тобто літератури, яка призначена для естетичного обслуговування освіченої частини суспільства з розвиненими культурними запитами [2].

Існує безліч жанрів та творів. У наш час йде тенденція комп'ютеризації усіх творів по жанрам. Актуальність цієї теми пов'язана з проблемою автоматизації цього процесу, яку може вирішити за допомогою спеціальних словників, що містять лексику певного жанру.

Метою цього дослідження є виділення формальних ознаки для ідентифікації жанрів й аналіз частотності використання лексичних одиниць, які ідентифікують зазначені жанри.

Матеріали і результати дослідження.

Літературні жанри (від фр. *genre* рід, вид) – групи літературних творів, що історично склалися, об'єднаних сукупністю формальних і змістовних властивостей (на відміну від літературних форм, виділення яких засноване тільки на формальних ознаках). Термін часто неправомірно ототожнюють з терміном «вид літератури» [2].

У якості одного із робочих визначень жанру ми притримувались визначення Р. Уеллека і О. Уоррена, які пропонують вважати жанром групу літературних творів, в яких теоретично виявляється загальна «зовнішня» (розмір, структура) та «внутрішня» (настрій, відношення, замисел, іншими словами – тема та аудиторія) форма [3].

Якщо на фольклорній стадії жанр визначався із внутрішньо-літературної (культурої) ситуації, то у літературі жанр отримує характеристику своєї сутності із власних літературних норм, кодифікованих риторикою. Уся номенклатура античних жанрів, що склалася до цього повороту, була потім енергійно переосмислена під його впливом.

Для кожного жанру масової культури створюється певний звід законів та правил – модель, яка забезпечує їх впізнання, тому читач ніколи не відчуває розчарування від нездійснених очікувань: у детективі злочинця завжди буде викрито, героїня жіночого роману знайде своє щастя, тощо. На відміну від елітарної культури масова культура говорить зі своїм читачем на зрозумілій йому мові [4].

Під час роботи проводився аналіз творів мовою оригіналу таких жанрів: детектив, наукова фантастика, фентезі, трилер та пригодницький роман.

Детектив (від англ. *detective* – слідчий) – один із найпродуктивніших жанрів масової культури ХХ століття, сюжетна колізія якого побудована на пошуку шляхів вирішення будь-якого складного і заплутаного завдання, в кінцевому підсумку пов'язана з набуттям істини. Центральною фігурою детектива є герой-одинак, в ідеалі – приватний детектив, що пов'язано зі специфікою знаходження гармонії у будь-якому тексті масової літератури, що будується на переконанні, що не державна система, а людина здатна змінити світ, зробити його краще. Засновниками жанру детектива є Е. По, У. Коллінз, А. Конан Дойл. Виділяють такі різновиди класичного детектива: англійська (аналітичний), де дія ослаблена, а сюжет складають

інтелектуальні логічні викладки геніального сищика (Шерлок Холмс А. Конан Дойля, Еркюль Пуаро А. Крісті); французький (екзистенціальний) детектив, де дія становить рефлексія сищика, який розкриває злочини, завдяки особливій душевній організації (герої Себастьяна Жапрізо); американський (жорсткий) детектив, де дія насичена подіями, головний герой – завжди активний, фізично сильна людина, здатний вийти із будь-якої ситуації (герої Д. Хемінґуей). В останні роки класичний детектив втрачає свої позиції, поступаючись місцем трилеру [4].

Трилер (від англ. *thrill* – тремтіти) – найбільш продуктивний жанр сучасного масового мистецтва, у другій половині ХХ століття практично повністю замінив собою класичний детектив, що пояснюється перш за все тим, що поняття істини, центральне для ідентифікації та самоідентифікації детектива, виявилися не таким істотним для світогляду людини епохи постсучасності.

Сюжет трилера, як правило, пов'язаний з ситуацією, коли у мирний перебіг життя персонажа раптово уривається якимось незрозумілим і неконтрольованим злом, з яким він змушений протистояти. Основною рисою трилера є така побудова сюжету, коли кожен наступний епізод пов'язаний із абсолютно несподіваним і, головне, практично не прогнозованим шляхом логічних побудов оборотом подій, «надзвичайною подією». При цьому межі епізодів також непередбачувані, що, у сукупності, викликає у читача специфічний стан напруженого «тривожного очікування». На відміну від детектива набуття істини у трилері з умовно детективним сюжетом є не наслідком продуманої стратегії і тактики слідчого, а, багато у чому, результатом збігу обставин, тобто ще одна несподіванка. Класичним взірцем трилера вважаються так звані «історії про маніяків», приклад чого – культовий текст кінця ХХ століття «Мовчання ягнят». Якщо у трилері тривожне очікування представляється як очікування неодмінно чогось жахливого, а події оточені містичним інфернальним ореолом, то подібна різновид жанру визначається терміном «чиллер». Найбільшим майстром, що працює у жанрі чиллеру, є Стивен Кінґ [4].

Фантастика (від грец. *φανταστική* – мистецтво уяви, фантазія) – жанр і творчий метод художньої літератури, кіно, образотворчого та в інших формах мистецтва, що характеризується використанням фантастичного допущення, «елемента надзвичайного», порушенням меж реальності, прийнятих умовностей. Сучасна фантастика включає в себе такі жанри як наукова фантастика, фентезі, жахи, магічний реалізм та багато інших [2].

Класифікація фантастики – це складний і суперечливий момент у фантастикознавстві. Загальноприйнята класифікація фантастики у даний час відсутня. Проблема класифікації пов'язана з нечіткістю самого терміна фантастика. Існує два підходи до визначення фантастики. Перший підхід полягає у тому, що фантастику визначають жанром

літератури і мистецтва. Другий підхід називає фантастику художнім прийомом в літературі і мистецтві. Ряд дослідників намагається об'єднати обидва підходи. Наприклад, Ольга Чигиринська в статті «фантастика: вибір жанру, вибір хронотопу» відзначає недоліки обох підходів і пропонує класифікувати фантастику по поєднанню специфічних хронотопом.

Один із поглядів на фантастику полягає у тому, що це література, яка доповнена фантастичним припущенням. На даний момент існують такі основні жанри фантастики наукова фантастика, фентезі, жахи. Часто також говорять про їх різновидах або «піджанрах» (таких, як альтернативна історія, кіберпанк або стімпанк). Фантастичні жанри, популярні в минулому, але рідко використовуються сучасними авторами – фантастична подорож і утопія [2].

Наукова фантастика (англ. скорочення sci-fi) (від грец. phantastikos – той, що стосується уяви) – жанр та метод у художній творчості; фантастика, в основі якої лежить екстраполяція на теми науки та технологій. В НФ-творах зображаються речі принципово можливі в рамках сучасної науки, засновані на теоріях або припущеннях. Важливе значення має також переконливе обґрунтування зображуваного. Досить часто, при вживанні терміну «фантастика», мається на увазі саме наукова фантастика. Засновником наукової фантастики вважається французький письменник Жюль Верн [3].

Особливістю творів наукової фантастики є те, що в них, як відомо, описується те, що ще не існує в реальній дійсності, але в принципі не суперечить законам її розвитку або ж може виникнути при певному збігу обставин. Наукової фантастику робить те, що ті чи інші її допущення або припущення побудовані на основі логічних висновків або з явищ сучасності, або з суми маловідомих фактів про минуле, які набувають в рамках художньої моделі імовірнісні ознаки (наприклад, польоти до далеких зірок або інопланетні цивілізації) [5].

Згідно з визначенням О. М. Осипова, наукова фантастика – це «літературний образний вислів наукових, соціальних та естетичних гіпотез і гіпотетичних ситуацій про минуле, сьогодення, які логічно проектується із явищ сучасності або сучасного світосприйняття, і тому імовірнісних або допустимих в рамках художнього експерименту, якими є твори» [5]. Наукова фантастика не займається прогнозуванням технічних чудес майбутнього, це не «література наукових ідей», сфера її застосування – дослідження життя соціуму та індивідуума, змін, що відбуваються в них під дією зрушень у розвитку науки і техніки. Завданням наукової фантастики у свій час вважалася популяризація досягнень науки і техніки, але ситуація змінилася в 60-х роках, коли до вивчення наукової фантастики звернулися не тільки літературознавці, а й філософи, історики, соціологи [5].

Найпомітнішими напрямками наукової фантастики є: альтернативна фантастика, соціальна фантастика, військова наукова фантастика, апокаліптична та

постапокаліптична фантастика, ксенофантастика, кіберпанк, космічна опера, феміністська наукова фантастика [5].

Фентезі – один із найпоширеніших жанрів сучасного формульного (масового) мистецтва. В основі фентезі лежить уявлення про існування паралельних світів, де розгортаються події твору [4].

Жанр фентезі за зовнішніми параметрами вважається різновидом фантастичної казки. Однак його відмінною рисою є те, що фентезійний світ не має нічого спільного зі звичною реальністю і населений вигаданими героями, які живуть згідно із законами свого світу, і лише психологія їх вчинків і дій може викликати традиційні, звичні асоціації у читача. В даному жанрі широко використовуються елементи ірреального, божественного [6].

Оскільки мова йде про інші реальності, практично в кожному тексті фентезі явно присутня міфологічна основа, яка орієнтована, як правило, на кельтську та давньоскандинавську міфологічні системи, на відміну від греко-римської міфології, сприймаються середньостатистичним читачем як «неосвоєна» територія. Текст фентезі дає можливість письменникові та читачеві сприймати те, що відбувається на абстрактному рівні, на рівні якоїсь моделі, що дозволяє гранично концептуалізувати погляди як з приводу основ світобудови, що конституюються, так і тих чи інших наболілих проблем сучасності.

Дж. Клют та Дж. Грант в «Енциклопедії фентезі» приводить до наступного визначення жанру: «Фентезійний текст – це внутрішньо узгоджена розповідь. Якщо дія відбувається у нашому світі, тоді історія, що викладається, у ньому неможлива (у нашому сприйнятті); якщо дія відбувається в іншому світі (otherworld), то неможливий сам цей світ, хоча історія у його рамках може бути цілком вірогідною. <...> фентезійний текст наполегливо запрошує читачів прожити (co-inhabit) історію разом з автором і героями. <...> Фентезі – це спосіб розповідати історії про фантастичне <...> на відміну від науково-фантастичних творів, які пишуться і читаються на основі припущення, що вони можливі – якщо ще не реальні» [6].

Е. Н. Ковтун у своєму навчальному посібнику «Художній вимисел у літературі ХХ століття» вводить термін *fantasy*, пояснюючи, що він «описує твори, в яких мотивація посилки, як правило, прибрана за межі тексту і, спираючись на принципи міфологічного мислення, конструє особливу модель світу», яку автор позначає як «справжню реальність» [7].

На підставі вищевикладеного вважається за доцільне в якості робочого визначення жанру фентезі вибрати досить ємне визначення В. Л. Гопмана, наведене в «Літературній енциклопедії термінів і понять»: «Фентезі – вид фантастичної літератури (або літератури про незвичайне), заснованої на сюжетному допущенні ірраціонального характеру. Це припущення не має «логічної» мотивації в тексті, припускаючи існування фактів і явищ, які чинять спротив раціональному поясненню» [2].

Практично всі дослідники жанру вважають, що передумови фентезі полягають в архаїчних міфах, середньовічних епосах і фольклорній чарівній казці. Значну роль у формуванні жанру відіграв художній досвід Середньовіччя та естетика лицарського роману, при цьому творчість найвидатніших письменників не тільки спиралась на архетипи й сюжети циклу про короля Артура, але й сама донині є джерелом архетипів [5].

Відомими представниками жанру фентезі є У. Ле Гуїн, Н. Перумов та ін. Елементи жанру фентезі присутні у романах Дж.К. Ролінг про Гаррі Поттера.

Пригодницький роман (також авантюрний роман, від фр. *aventure*) – жанр роману, що сформувався в середині XIX століття на хвилі романтизму і неоромантизму з характерним для них прагненням бігти від міщанської повсякденності у світ екзотики і героїзму. Це один з традиційних жанрів масової літератури.

У більш широкому сенсі можна говорити про існування особливого авантюрного жанру, або пригодницької літератури, яку розрізняють різкий поділ персонажів на героїв і лиходіїв, «стрімкість розвитку дії, мінливість і гострота сюжетних ситуацій, перебільшеність переживань, мотиви викрадення і переслідування, таємниці і загадки» [1].

Авантюрний (пригодницький) роман є однією з найбільш вільних жанрових форм, яка увібрала у себе риси інших композиційно-стильових явищ. Поява авантюрного роману пов'язано з утвердженням принципу вільної поведінки людини і її самодостатності. Родоначальником авантюрного роману нерідко називають А. Дюма-батька.

Серед ознак авантюрного роману виділяються насиченість сюжету, у якому події стрімко змінюють одна одну, чітка композиція з зав'язкою, кульмінацією і розв'язкою, пріоритет дії над психологічним аналізом, чітка поляризація сил добра і зла.

Авантюрний роман створював модель «другої реальності», для нього характерні містифікація, пародіювання, іронія, що дозволяли вийти з лещат поступово затверджувалася цензури [8].

Завдання пригодницької літератури не стільки повчати, аналізувати або описувати реальність, скільки розважати читача [1].

Попередниками авантюрного роману XIX століття були романи крутійські, готичні та історичні (Вальтер Скотт, Фенімор Купер, Віктор Гюго). Помітний внесок у становлення жанру внесли такі майстри романтичної і неоромантичної прози, як Олександр Дюма у Франції і Р. Л. Стівенсон в Великобританії. Спочатку розрізнялися романи історико-пригодницькі («Три мушкетери», «Королева Марго») і соціально-пригодницькі, з сучасного життя («Граф Монте-Крісто», «Паризькі таємниці»).

У масовій літературі жанри різняться за тематикою, а тематика визначається певною лексикою. Для пригодницького роману було обрано тематику морських пригод, для жанру наукової фантастики – космічних подорожей.

Частотний словник (або частотний список) – набір слів даної мови (або підмови) разом з інформацією про частоту їх зустрічальності.

Частотні списки використовуються для викладання мови, створення нових словників, додатків комп'ютерної лінгвістики, досліджень в області лінгвістичної типології, тощо [9].

Частотна система пристрою дозволяє визначити, які з його одиниць найбільш і менш уживані, рідкісні, тобто по заданому рангу або частоті знаходити відповідні їм одиниці словника. Однак, помилково називати частотним тільки словник із розташуванням його матеріалу по частоті [9]. Частотна кваліфікація словникових одиниць не тільки не заперечує, але і передбачає алфавітно-частотне оформлення словника поряд із власне частотним. Алфавітний варіант необхідний для визначення частот (і далі рангів) заданих словникових одиниць, що важливо в навчальних цілях і при типологічних дослідженнях різних текстових сукупностей та текстів [9].

Поняття «частотний словник», таким чином, охоплює не стільки систему розміщення словникового матеріалу, скільки принцип його відбору (у такій словник потрапляють тільки ті одиниці, які зустрілися у порядку в проаналізованому тексті) та чисельну реєстрацію його фактичного споживання. Тому частотний словник може бути влаштований по кожній із трьох основних словникових систем, тобто крім уже реалізованих алфавітній (алфавітно-частотної) і частотної систем, можливо також побудова асоціативних (картинних, тематичних і навіть тезаурусів) частотних словників.

Один з перших частотних словників, словник китайських ієрогліфів, складався для оптимального пристрою набраних кас в друкарні. Найбільший частотний словник, другий з існуючих за обсягом вибірки і опублікованого словника, створювався на матеріалі німецьких текстів з метою удосконалення системи стенографії. Більшість зарубіжних частотних словників мали на меті забезпечити методику навчання мови (навички читання і письма) тих, для кого ця мова є нерідною. Менша кількість частотних словників, переважно в останні роки, призначена для лінгвістичних і лінгвостилістичних досліджень. Відомо невелике число словників, розрахованих на технічні, лінгвопсихологічні та інші спеціальні, наприклад, військово-технічні та юридичні додатки.

Особливий інтерес представляють статистичні дослідження дитячого мовлення, зокрема, вивчення динаміки розвитку словника дитини. Такі роботи дозволяють робити спостереження, однаково важливі для лінгвістів і психологів [9].

Аналіз частот лінгвістичних одиниць у мовленні з патологічними відхиленнями дозволяє судити про наявність та стадіях хворобливого процесу у осіб, схильних до психічних захворювань.

За останні роки роботи зі створення частотних словників на базі вибірок обсягом від 500 тис. до 1 млн. зробили слововживань і трохи вище мають на меті отримати матеріалу для різнобічного

лінгвістичного аналізу тексту на рівнях словоформ, слів, морфем, графем, графемо. Автори цих словників виробляють не тільки лексикологічний аналіз тексту, а й досліджує його на інших лінгвістичних рівнях; частотний словник, таким чином, є для них лише початковим етапом ретельного вивчення структури тексту, що представляє «мова в цілому».

Частотні словники групи «Статистика мови» є багатоцільовими. Вони розраховані на забезпечення лексичним і лексико-морфологічним мінімумом викладачів іноземних мов і, перш за все, викладачів немовних навчальних закладів, особливо гостро відчують необхідність у підвищенні ефективності своєї роботи. Вони призначені також для використання при створенні систем автоматичної переробки мовної інформації. Вони дають матеріал для дослідження типології функціональних стилів і мов, так як складаються за єдиною методикою і базуються зазвичай на вибірках стандартного обсягу. Ці дві властивості – єдність методики та однаковий обсяг вибірки – мають принципове значення для можливості порівнювати дані частотних словників [9].

Особливе місце серед частотних словників займають словники, які можуть бути використані в автоматичних інформаційно-пошукових системах. Так, «Частотний словник індексування» (під редакцією Л. В. Цукрового) складено для вирішення інформаційно-пошукових завдань. Матеріалом для словника послужили 1660 рефератів з електровимірювальні прилади (всього близько 105 тисяч слововживань). Принципи складання словника застосовні до будь-якої галузі знань. В основу словника покладено нову для лексикографії одиницю – гіперлексему (безліч спільнокоренових лексем однієї або різних частин мови, пов'язаних між собою відносинами трансформації, похідне: інерція, інерційний, інерційність; зуб, зубець, зубчастий, зубцову. У словнику гіперлексеми розміщені за ступенем частотності [9].

Відомі такі частотні словники: Yosselson H. The Russian Word Count and Frequency Analysis of Grammatical Categories of Standard Literary Russian. – Detroit, 1953; Штейнфельдт Э. А. «Частотный словарь современного русского литературного языка». «Частотный словарь русского языка» за редакцією Л. Н. Засориної. В Україні в 1981 р. вийшов двотомний «Частотний словник сучасної української художньої прози».

Дослідження частотності лексичних одиниць дозволяє зробити важливі висновки про особливості мови і стилю письменника. Так, «Частотний словник роману Л. Н. Толстого «Війна і мир» складається з алфавітного списку всіх слів, які розміщені в порядку убавання частот, і статистичної структури словника, що представляє собою таблицю розподілу частот. Словник відкриває широкі можливості для наступного аналізу мови письменника, виявлення статистичних особливостей його стилю. В примітках коротко пояснюються спеціальних, архаїчних і діалектних слів [9].

«Частотний словник оповідань А. П. Чехова»

містить упорядкований по частотах перелік 13736 лексем, що налічуються в 150 оповіданнях А. П. Чехова. До словника додається перелік назв (874 лексеми) і таблиця їх частотного розподілу. Словник є першим в рамках статистикою – лексикографічного проекту «Лінда» [9].

Зазвичай частотні словники будуються на основі корпусів текстів: береться набір текстів, який є представником для мови в цілому, для деякої предметної області або даного автора і з нього витягуються словоформи, леми і частини мови (останні витягуються у разі, якщо корпус має морфологічну розмітку).

Проблеми при створенні частотних списків полягають у: відтворюваності (чи будуть результати ідентичні на іншому аналогічному корпусі); сплесках частоти окремих слів (частота слова в одному тексті може вплинути на його позицію в частотному списку); складності визначення позиції менш частотних слів, що не дає можливості ранжувати їх раціонально.

Усі ці проблеми пов'язані з тим, що зі статистичної точки зору мова являє собою велику кількість рідкісних подій (Закон Ціпфа), в результаті чого невелика кількість слів зустрічається дуже часто, а переважна більшість слів мають дуже невисоку частоту.

Для опису сплесків частоти можна використовувати метафору хоббіта (Адам Кілгарріфф спочатку використовував відносно рідкісне англійське слово *whelk*, вид морського моллюска): якщо кілька текстів у корпусі про хоббітів, то це слово буде вживатися майже у кожному реченні. У результаті його частота у цих текстах буде порівнянна із частотою службових слів, але у частотному списку великого корпусу, в який входять такі тексти, це слово буде мати неправдоподібно високий ранг. Такі сплески частоти можна оцінювати за допомогою коефіцієнта варіації: відносини стандартного відхилення до середньої частоти.

Частотні словники забезпечують можливість порівняти два корпуси, щоб визначити слова, найбільш характерні для кожного з них. У зв'язку з тим, що розміри корпусів можуть бути різні, більш надійна оцінка частоти слів ґрунтується на приведенні їх до ЧМС (частота на мільйон словоформ, англ. *Ipm, instances per million words*). Список літератури:

Для визначення набору ключових слів, що відрізняють один корпус від іншого можна використовувати різні статистичні заходи: χ^2 -квадрат, відношення правдоподібності (англ. *Likelihood-ratio test*), тощо [9].

Для рішення завдання автоматизації створення частотних словників для жанрів масової літератури було створено наступний алгоритм:

До програми підключаються жанрові словники створені авторами заздалегіть (імітування інтелектуального навчання системи): детектив, наукова фантастика, фентезі та пригодницький роман. Попередньо вони були створені експертом під час виділення жанро-визначальної лексики. Також до

програми підключається словник загальної лексики та список стоп-слів.

На вхід програми подається текст. Програма виключає стоп-слова. Проводиться лематизація та морфологічний аналіз. Виключаються усі частини мови, окрім іменників і рахується їхня частота.

Програма порівнює отриманий список іменників

із словником експерта. Якщо слова немає у словнику експерта, тоді програма відносить це слово до словника певного жанру. Якщо слово є у словнику експерта, тоді програма виключає це слово із списку іменників з частотою із вхідного тексту. Приклад роботи програми наведено на рис. 1.

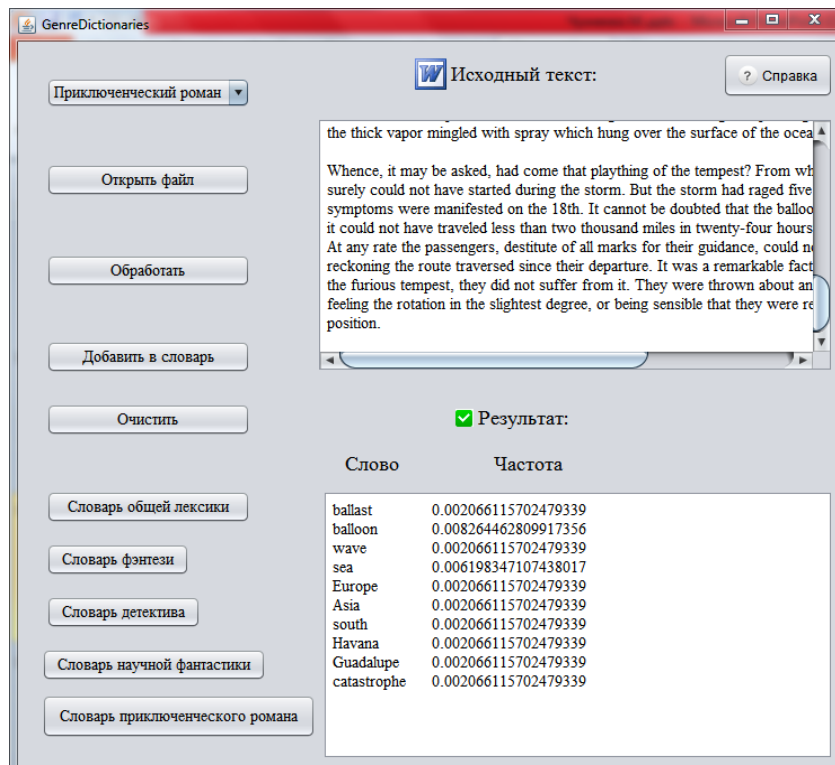


Рис. 1 – Результат роботи програми автоматичного визначення ключових слів

Висновки. У результаті роботи програми отримуємо частотні словники чотирьох жанрів, для кожного з яких буде свій невеликий частотний словник, у якому зібрані найхарактерніші слова конкретного жанру. Наявність таких словників дозволить в подальшому автоматизувати процес визначення жанру для будь-якого твору масової літератури.

Список літератури: 1. *Купина Н. А.* Массовая литература сегодня / *Н. А. Купина, М. А. Литовская, Н. А. Николкина.* – М.: Флинта-Наука, 2009. – 424 с. 2. *Николюкин А. Н.* Литературная энциклопедия терминов и понятий / *А. Н. Николюкин.* – М.: Интелвак, 2001. – 1600 с. 3. *Уэллек Р.* Теория литературы / *Р. Уэллек, О. Уоррен.* – М.: Прогресс, 1978. – 325 с. 4. *Суслова Н. В.* Новейший литературоведческий словарь-справочник / *Н. В. Суслова, Т. Н. Усольцева.* – Мозырь: Белый ветер, 2003. – 43 с. 5. *Осипов А. Н.* Фантастика от "А" до "Я" (Основные понятия и термины) / *А. Н. Осипов.* – М.: Дограф, 1999. – 752 с. 6. *Clute J.* The Encyclopedia of Fantasy / *J. Clute, J. Grant.* – London: Orbit, 1999. – 1076 p. 7. *Ковтун Е. Н.* Художественный вымысел в литературе XX в. / *Е. Н. Ковтун.* – М.: Высш. шк., 2008. – 406 с. 8. *Черняк В. Д.* Массовая литература в понятиях и терминах / *В. Д. Черняк,*

М. А. Черняк. — М.: ФЛИНТА, 2015. — 192 с. 9. *Алексеев П. М.* Частотные словари / *П. М. Алексеев.* – Санкт-Петербург: Изд-во С.-Петерб. ун-та, 2001 – 156 с.

Bibliography (transliterated): 1. *Kupina N. A., Litovskaya M. A., Nikolina N. A.* *Massovaya literatura segodnya* [Mass literature today]. Moscow, Flinta-Nauka, 2009. 424 p. 2. *Nikolyukin A. N.* *Literaturnaya entsiklopediya terminov i ponyaty* [Literary Encyclopedia of Terms and Concepts]. Moscow, Intelvak, 2001. 1600 p. 3. *Uellek R.* *Teoriya literature* [Theory of Literature]. Moscow, Progress, 1978. 325 p. 4. *Suslova N. V., Usoltseva T. N.* *Noveyshy literaturovedchesky slovar-spravochnik* [The newest literary reference dictionary]. Mozyr, Bely veter, 2003. 43 p. 5. *Osipov A. N.* *Fantastika ot "A" do "Ya" (Osnovnye ponyatiya i terminy)* [Sci-Fi from "A" to "Z" (Basic concepts and terms)]. Moscow, Dograf, 1999. 752 p. 6. *Clute J.* *The Encyclopedia of Fantasy*. London, Orbit, 1999. 1076 p. 7. *Kovtun Ye. N.* *Khudozhestvenny vymysel v literature XX v.* [Art fiction in the literature of the twentieth century]. Moscow, Vyssh. shk., 2008. 406 p. 8. *Chernyak V. D., Chernyak M. A.* *Massovaya literatura v ponyatiyakh i terminakh* [Mass literature in terms and terms]. Moscow, FLINTA, 2015. 192 p. 9. *Alekseyev P. M.* *Chastotnye slovari* [Frequency dictionaries]. Saint-Petersburg, Izd-vo S.-Peterb. un-ta, 2001 156 p.

Надійшла (received) 17.06.2017

Бібліографічні описи / Библиографические описания / Bibliographic descriptions

Особливості класифікації творів масової літератури / Н. С. Угольнікова, М. В. Чухненко // Вісник НТУ «ХП». Серія: Актуальні проблеми розвитку українського суспільства. – Харків: НТУ «ХП», 2017. – № 29 (1251). – С. 91–97. – Бібліогр.: 9 назв. – ISSN 2227-6890.

Особенности классификации произведений массовой литературы / Н. С. Угольникова, М. В. Чухненко // Вісник НТУ «ХПІ». Серія: Актуальні проблеми розвитку українського суспільства. – Харків: НТУ «ХПІ», 2017. – № 29 (1251). – С. 91–97. – Бібліогр.: 9 назв. – ISSN 2227-6890.

Features of the classification of works of mass literature / N.S. Uholnikova, M.V. Chukhnenko // Bulletin of NTU "KhPI". Series: Actual problems of Ukrainian society development. – Kharkiv: NTU "KhPI", 2017. – № 29 (1251). – P. 91–97. – Bibliogr.: 9. – ISSN 2227-6890.

Відомості про авторів / Сведения об авторах / About the Authors

Угольникова Наталія Сергіївна – кандидат філологічних наук, Національний технічний університет «Харківський політехнічний інститут», доцент кафедри інтелектуальних комп'ютерних систем; тел.: (050) 576–43–56; e-mail: natalie.uns@gmail.com.

Угольникова Наталия Сергеевна – кандидат филологических наук, Национальный технический университет «Харьковский политехнический институт», доцент кафедры интеллектуальных компьютерных систем; тел.: (050) 576–43–56; e-mail: natalie.uns@gmail.com.

Uholnikova Natalia Serhiivna – Candidate of Philological Sciences (Ph. D.), National Technical University "Kharkiv Polytechnic Institute", Associate Professor of the Intelligent Computer Systems Department; tel.: (067) 72–65–111; e-mail: natalie.uns@gmail.com.

Чухненко Маргарита Вячеславівна – Національний технічний університет «Харківський політехнічний інститут», студентка; тел.: (093) 664–67–34; e-mail: rita18_1995@mail.ru.

Чухненко Маргарита Вячеславовна – Национальный технический университет «Харьковский политехнический институт», студентка; тел.: (093) 664–67–34; e-mail: rita18_1995@mail.ru.

Chukhnenko Marharyta Viacheslavivna – National Technical University "Kharkiv Polytechnic Institute", student; tel.: (093) 664–67–34; e-mail: rita18_1995@mail.ru.