

Использование информационно-поисковой системы Интегрум в гуманитарных исследованиях

Using «Integrum» Information Retrieval System in Human Studies

Використання інформаційно-пошукової системи Інтегрум у гуманітарних дослідженнях

Е. В. Новицкая

Представительство компании MIPP International, Минск, Беларусь

Ekaterina Novitskaya

MIPP International Representative Office, Minsk, Belarus

К. В. Новицька

Представництво компанії MIPP International, Мінськ, Білорусь

Рассматриваются возможности библиотек и исследователей при работе с базами данных Интегрум, крупнейшим в мире электронным архивом СМИ России и стран СНГ, объединяющим в себе более 7000 баз данных. Помимо качественно нового информационного пространства Интегрум предлагает уникальную совершенную систему поиска и удобные аналитические инструменты для обработки данных. Использование Интегрума в гуманитарных исследованиях демонстрируется на примере изучения распространения художественного жанра «фэнтези» в России.

The potential of Integrum databases, the largest archive of Russian and CIS countries' mass media, for using in libraries and research is examined. Integrum incorporates over 7,000 databases. Beside the new information space, Integrum offers the unique and perfect retrieval system, as well as friendly analytical tools for data processing. Using Integrum in human research is illustrated by the example of the study of the fantasy genre in Russian literature.

Розглянуто можливості бібліотек та дослідників під час роботи із базами даних Інтегрум, найбільшим у світі електронним архівом ЗМІ Росії та країн СНД, що об'єднує понад 7000 баз даних. Разом із якісно новим інформаційним простором, Інтегрум пропонує винятково досконалу систему пошуку та зручні аналітичні інструменти для обробки даних. Використання Інтегрум у гуманітарних дослідженнях продемонстровано на прикладі вивчення розповсюдження художнього жанру «фентезі» у Росії.

Компания MIPP International – это всемирно известный поставщик книжной продукции из России и стран СНГ. Имея немалый опыт работы с библиотеками по всему миру, мы знаем, что с каждым днем возрастает интерес к информации в электронном виде. В 2007 году мы стали партнерами с компанией Integrum World Wide и, идя в ногу со временем, начали заниматься совершенно новым для нас видом деятельности – предоставлением доступа к базам данных Интегрум библиотекам за пределами РФ.

Информационно-поисковая система Интегрум вызывает во всем мире огромный интерес исследователей. Она посвящена современной России, различным сторонам ее жизни и их отражению в текущей печати. Мы называем Интегрум электронным портретом России и стран СНГ, ведь ни одна другая база данных не содержит в себе такое огромное количество полнотекстовых документов, касающихся стран бывшего Советского Союза – более 360 000 000 документов. Количество источников информации уже превысило 7 000, и это далеко не предел, ведь ежедневно эта коллекция пополняется новыми источниками информации, а это в среднем 40 000 новых документов в день.

В первую очередь Интегрум интересен нашим подписчикам как крупнейшая электронная коллекция средств массовой информации России и стран СНГ. По данным на 17 апреля 2008 года базы данных Интегрума содержали электронные версии 5836 источников СМИ. (см. Таб. 1).

СМИ: Центральные газеты	378
СМИ: Центральные журналы	747
СМИ: Региональная пресса	2171
СМИ: Мировые информационные агентства	120
СМИ: Информагентства РФ, СНГ.	478
СМИ: Мониторинг	234
СМИ: Теле-Радио Эфир	182
СМИ: Зарубежная пресса	287
СМИ: Интернет-издания	1170
СМИ: Компьютерная пресса	69

Таблица 1. Базы данных средств массовой информации, представленные в Интегруме (по данным на 17 апреля 2008 года).

Группа баз данных «Библиотечные фонды» включает в себя документы Роспатента с 1991 года по настоящее время, карточки библиотеки Агропрома, ИНИОН, Российского Государственного Архива Литературы и Искусства (РГАЛИ), Российской книжной палаты (РКП), Государственной публичной научно-технической библиотеки России (ГПНТБ)

В группе «Адресно-справочные базы данных» содержится 160 источников информации, среди которых Большой энциклопедический словарь, Большая Советская Энциклопедия, справочник «Кино российское и советское», регистр предприятий России и Украины и многое другое.

Вниманию наших подписчиков также представлены полные тексты тысяч произведений русской классики – собрания сочинений Пушкина, Чехова, Цветаевой, выдающиеся произведения Толстого, Достоевского, Мандельштама, Ахматовой, Набокова, Булгакова, Ильфа и Петрова, и т. д. Кроме этого, имеется возможность прослушать тексты многих произведений в исполнении профессиональных актеров. Для зарубежных исследователей интерес представляет то, что многие произведения представлены и на русском и на английском языке параллельно.

Кроме всего этого, Интегрум располагает документами законодательства России, биографическими, страноведческими данными. Многих библиотекарей интересуют каталоги промышленной продукции, информация официальных учреждений, статистические документы (Госкомстата, ГМЦ Росстата, РЦЭР, Статкомитета СНГ).

Документы в базах данных Интегрум представлены на русском, украинском и английском языках. Кроме этого, любой документ может быть автоматически переведен с русского на английский или немецкий язык.

Как вы понимаете, поиск среди такого количества документов – задача не простая. Поэтому создатели Интегрума постарались как можно больше облегчить ее для пользователей. Поиск проводится по всему массиву баз данных одновременно и выдает результаты за считанные секунды. Система ищет все словоформы запроса, учитывает возможные опечатки, проверяет слово по нескольким словарям одновременно: орфографическом, толково-словообразовательном, словаре трудностей, синонимов, антонимов, ударений, личных имен.

Первоначально Интегрум не создавался для ученых, основными его пользователями были бизнесмены и журналисты, для которых Интегрум являлся, главным образом, удобной библиотекой, в которой можно было быстро найти любую необходимую информацию. Эти пользователи сразу оценили преимущества Интегрума по сравнению с Интернетом. Ведь Интернет не гарантирует качество информации, там отсутствуют полнотекстовые архивы средств массовой информации, а его поисковые машины не имеют настолько гибкого языка запросов для точного поиска.

Интегрум предложил качественно новое поисковое пространство и новые средства самого поиска. Эти достоинства Интегрума очень быстро привлекли к нему ученых, прежде всего политологов, социологов и филологов.

Первые попытки применения Интегрума в гуманитарной сфере ограничивались сбором примеров. С его помощью стало возможным доказать гуманитарную гипотезу на основе тысяч и десятков тысяч примеров. Это стало немалым достижением, так как придавало доказательству репрезентативность, строгость и точность. Однако этого было недостаточно. Когда в Интегруме были

разработаны сервисы «Сравнительная и относительная упоминаемость» и «Частотный словарь русского языка», ученые–лингвисты окончательно убедились в полезности практического применения Интегрума.

В качестве примера я бы хотела исследовать с помощью Интегрума активное распространение в России художественного жанра фэнтези. Его родоначальником в мире считается известный английский писатель Джон Рональд Руэл Толкиен. Его книга «Хоббит» и трилогия «Властелин колец», написанные в середине 20 века, были переведены на русский язык в 80-х. В начале 90-х интерес к фэнтези-литературе в России многократно возрастает. Первое упоминание слова фэнтези в базах Интегрума относится к 1990 году:

«... Рассказ «Охотник на ягуаров» прекрасно демонстрирует особенности художественного метода Шепарда. Время действия у него, как правило, — настоящее или близкое будущее. Место — чаще всего — берега Карибского моря или Мексиканского залива. Жанр — что-то среднее между научной фантастикой и **фэнтези**.» (Вокруг света; 15. 09. 1990; 9)

Начиная с этого момента, динамику частотности употребления слова «фэнтези» в центральной и региональной российской прессе Интегрум демонстрирует следующим образом (см. Рис. 1):



Рис. 1. Динамика частотности упоминаемости слова «фэнтези» в Центральной и региональной прессе России

С появлением и популяризацией фэнтези в России в русском языке получили распространение многие слова иностранного происхождения, ранее нечасто использовавшиеся. Такие мифонимы как «эльф», «орк», «тролль», «гном» и т. п. становятся все более и более популярными. Следующий график (Рис. 2) демонстрирует использование вышеуказанных мифонимов в прессе России с 2001 по 2004 год.



Рис. 2. Динамика частотности упоминаемости мифонимов, использующихся в книгах жанра «фэнтези», в российской прессе

Таблица 2 показывает точное количество упоминаний мифонимов жанра фэнтези в российской прессе с июня 2001 года по апрель 2008 года.

30. 6. 2001	30. 9. 2001	31. 12. 2001	31. 3. 2002	30. 6. 2002	30. 9. 2002	31. 12. 2002
875	590	958	990	834	892	1165
31. 3. 2003	30. 6. 2003	30. 9. 2003	31. 12. 2003	31. 3. 2004	30. 6. 2004	30. 9. 2004
1173	1099	1063	1575	1720	1707	1507
31. 12. 2004	31. 3. 2005	30. 6. 2005	30. 9. 2005	31. 12. 2005	31. 3. 2006	30. 6. 2006
1958	1502	1970	1495	1819	1411	1620
30. 9. 2006	31. 12. 2006	31. 3. 2007	30. 6. 2007	30. 9. 2007	31. 12. 2007	31. 3. 2008
1731	2022	1595	1618	1950	2327	2147

Таблица 2. Точное количество упоминаний мифонимов из произведений литературы «фэнтези» в российской прессе с 2001 по 2008 год (ежеквартальные показатели).

Любопытно, что пик использования данных слов приходится на моменты выхода экранизации трилогии Толкиена «Властелин колец». 7 февраля 2002 года в России выходит первый фильм и количество упоминаний мифонимов достигает 990, а это в два раза больше, чем 6 месяцами ранее. 23 января 2003 года состоялась всероссийская премьера второго фильма, и количество упоминаний вырастает почти на 2 сотни, а после выхода третьей части, в декабре 2003 года – 1575 упоминаний.

Вот лишь некоторые из цитат, найденных с помощью Интегрума с 2001 по 2004 год:

«... Театральный **орк** является одним из самых неприятных представителей сценического Средиземья. Орк – это **эльф**, у которого не задалась карьера, и он остановился в развитии. « (Известия; 13. 02. 2002)

«... Им противостоят **орки, тролли**, нежить и другая часть людей... Книги жанра «фэнтези» (фантастика с уклоном в волшебство и Средневековье) по продажам лишь немного уступают вездесущим детективам в мягких обложках – причем произведения не зарубежных, а наших авторов... Появился даже жанр «славяно-киевской фэнтези», использующий языческие легенды Древней Руси и вообще языческую эстетику. « (Парламентская газета; 22. 08. 2003)

Однако, даже по прошествии времени данная лексика не перестает быть популярной. Это связано в первую очередь с тем, что огромное количество писателей с головой погружается в новый жанр. Фэнтези набирает почитателей ежедневно, становясь один из самых популярных литературных жанров. Во-вторых, увлечение новым жанром фэнтези влечет за собой развитие ролевого движения. На западе оно прижилось уже давно, а вот в Москве и крупных городах России его первый всплеск приходится на конец 80-х. В начале 21 века все больше и больше людей стремятся окунуться в мир эльфов, гномов и хоббитов, а средства массовой информации все чаще посвящают им свои статьи:

«.. Лично мне врезалась в память информация о том, что у нас в стране *внезапно появились сказочные жители: эльфы, гномы, гоблины и прочая...* Если вам все-таки повезет и вы увидите кого-либо из этих загадочных существ, прошу: не пугайтесь, не сердитесь, а просто полюбуйтесь на них. Ведь пока они существуют, будет жить и волшебство, по крайней мере – в нас самих» (Советский Сахалин (Южно-Сахалинск); 02. 04. 2004; 57)

«.. Потом собираются в московских перелесках, где участвуют в ролевых играх: называют друг друга *эльфийскими* именами, учат тайный *гномий* язык, дерутся на самодельных деревянных мечах, ищут кольца всевластия...» (Меридиан КМВ (Пятигорск); 23. 11. 2004; 47 (71))

«.. Красногорский р-н, оздоровительный центр «Зоркий». Здесь народ ориентирован на модный нынче стиль фэнтези. Аня, воспитатель: «*Вожатых мы называем магами, начальство лагеря – верховными магами... Всего в лагере семь племен, среди которых – муми-тролли, эльфы, дриады и гномы*» « (Аргументы и факты; 03. 07. 2002; 27)

Еще один любопытный факт. Книга «Властелин колец» по данным на начало апреля 2008 года упоминалась в российской прессе 8307 раз. Однако, как было замечено в последнее время, все чаще и чаще словосочетание «властелин колец» употребляется в совершенно новом контексте, например, в статьях на спортивную тему:

«... «Гвоздём» праздника стала спортивная эстафета для школьников «**Властелин колец**» (каждая школа была представлена сборной командой в составе пяти человек). « (Вечерний Томск; 04. 12. 2007; 226)

«... В Вологде стартовали соревнования по баскетболу «**Властелин кольца**», посвященные 860-летию города Вологды... Организаторами «Властелина кольца» выступают городская федерация баскетбола и управление физической культуры, массового спорта и молодежи администрации города. « (Вологодские новости (Вологда); 14. 06. 2007; 023)

Инструмент «Относительная статистика» позволяет отследить, как развивается употребление данного словосочетания в спортивном контексте. График (Рис. 3) демонстрирует, что именно региональная пресса стала первой ассоциировать его со спортом.

Однако, употребление словосочетания «властелин колец» не ограничивается спортивной темой. Его использование при упоминании марки автомобиля «Ауди» представлено в 93 документах. В первую очередь, властелином колец называют сам автомобиль:

«... Хорошая новость для поклонников «*властелина колец*». Тем, кому обычное купе *Audi A5* вдруг покажется недостаточно динамичным, спортивное подразделение немецкой компании *Quattro GmbH* предлагает два собственных пакета «переделок» *S-Line*. « (Автобизнес-Weekly (Минск); 26. 04. 2007; 17)

«... *Audi A4 – властелин колец в бизнес-классе*» (Деловая неделя (Киев); 27. 09. 2007 21:59)

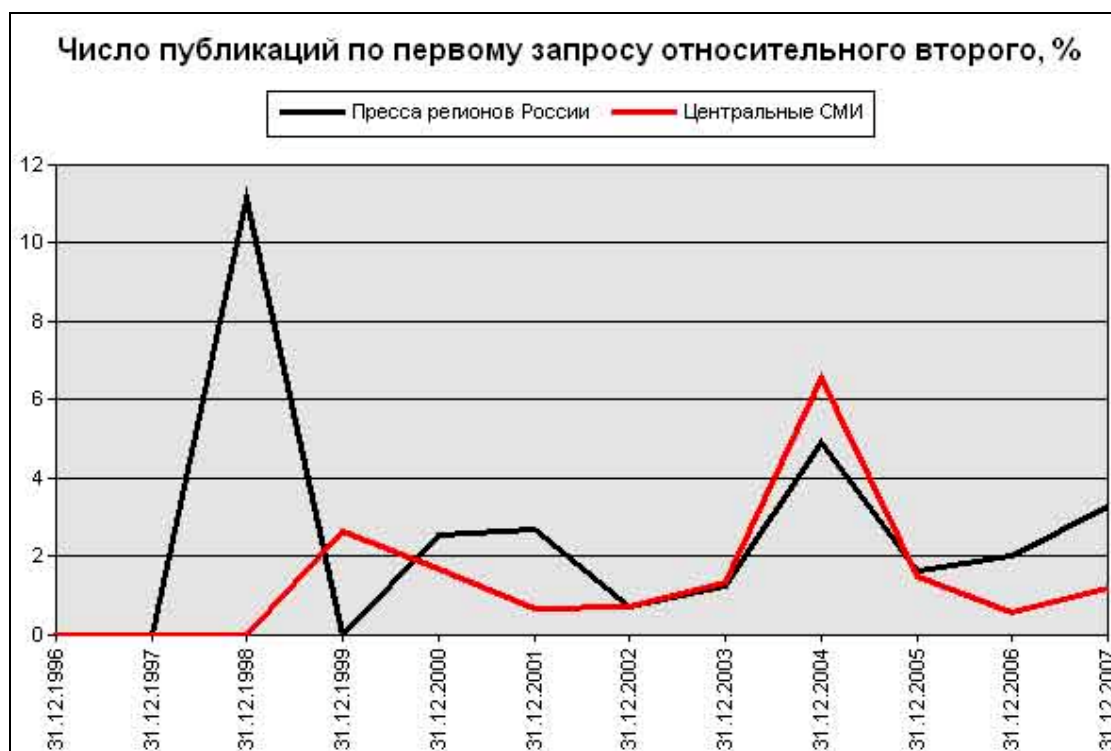


Рис. 3. Употребление словосочетания «властелин колец» в контексте спортивных событий

Во-вторых, так нередко называют самих владельцев автомобиля «Ауди» :

«... И хотя в народе обладателей автомобиля марки Audi из-за эмблемы машины, на которой четыре кольца соединены между собой, уже давно в шутку называют **властелинами колец**, Сафаров больше всех заслуживает такого прозвища. « (Корреспондент (Киев) 44; 11. 11. 2006)

«... Мэр Красноярска Петр Пимашков благодаря своему автомобилю получил прозвище «**властелин колец**». Его машина – серебристая Audi с номером «А 001 00» « (Новые известия (Москва); 25. 08. 2006; 152)

Как мы видим, фэнтези уже не только прочно вошло в нашу жизнь как новый литературный жанр, оно постепенно вливается во все сферы нашей жизни, и даже вносит свой вклад в появление новых слов и их неожиданных значений в русском языке. Это небольшое гуманитарное исследование могло бы продолжаться бесконечно. Интегрум поможет узнать предпочтения российской прессы в написании фамилии автора «Властелина колец» : (Толкиен или Толкин?), в написании самого названия жанра («фэнтези» или «фэнтэзи?»), узнать о развитии «славянского фэнтези» и других разновидностей жанра.

Возможности применения Интегрума в гуманитарных исследованиях действительно очень широки. Многие ученые считают, что эта информационная система является сокровищем для специалиста, изучающего современную Россию и русский язык. Самое главное достоинство Интегрума на их взгляд это то, что с помощью этой системы можно выполнять работы, основывающиеся не на предположениях или случайных наблюдениях, а на конкретных фактах и примерах.

Таким образом, система Интегрум создает для библиотек ряд новых возможностей, позволяющих им в сложных современных условиях добиваться максимально возможного уровня информационного обеспечения науки, невозможного при использовании только традиционных научных изданий.

Литература

1. Никипорец-Такигава Г. (2006): *Integrum: точные методы и гуманитарные науки*. Летний Сад, Москва.
2. Материалы баз данных СМИ России и стран СНГ «Интегрум».