

# **БІБЛІОТЕКА ІНФОРМАЦІЙНОГО СУСПІЛЬСТВА: ЦІННІСНІ ОРІЄНТИРИ ДІЯЛЬНОСТІ**

УДК 004.6 : 024.4

## **ИСПОЛЬЗОВАНИЕ АВТОРИТЕТНОГО ФАЙЛА АВТОРОВ В ПОИСКЕ В АРМЕ ВЕБ-ИРБИС**

**Панарин Г. С.,** ведущий библиограф

**Научная библиотека Харьковского  
национального технического университета сельского  
хозяйства им. П. Василенко**

Вопрос лингвистического обеспечение автоматизированных библиотечно-информационных систем (АБИС) традиционно поднимается на многих конференциях и семинарах. На данный момент озвучена и продемонстрирована масса примеров механизмов унификации поисковых терминов и ведения поискового человеко-машинного диалога с использованием информационно-поисковых языков (ИПЯ). В данном докладе хотелось бы еще раз затронуть тему Авторитетных файлов (АФ), механизмов их создания и использования при поиске.

АФ – это средство унификации информации, используемой для каталогизации и поиска. И если функция каталогизации с помощью АФ сводится именно к вводу нормированного термина в правильной форме, то функции поиска с использованием АФ гораздо шире. Частным случаем АФ является тезаурус, в котором к каждому термину (дескриптору) создается поисковый образ, который может включать перечень синонимов, надлежащих и подлежащих рубрик, связанных индексов, тематических заголовков – всего того, что может быть связано с данным термином и использовано в качестве

поисковых элементов в запросах. И именно этому будет посвящен мой доклад.

Особенностью формирования поисковых словарей в АБИС Ирбис является то, что словари формируются полностью по самим введенным данным. Другими словами, как введено, так и отображается. В некоторых случаях это создает определенные проблемы. Так в случае с авторами, которые на документе могут указываться во множестве разных форм, в словаре формируется на одно лицом множество поисковых терминов. А если учесть, что у автора могут быть статьи изданы на разных языках, то количество форм его имени растет и вовсе в геометрической прогрессии. Это приводит к тому, что в поисковый запрос приходится отбирать несколько терминов, описывающих одного и того же человека. Именно с такой проблемой столкнулись сотрудники нашей библиотеки, когда начали работу по подбору литературы своих авторов сотрудников. Желание сломать сложившуюся ситуацию и создать возможность получения полной информации о работах того или иного автора с как можно меньшими затратами привело к возникновению идеи подключения для поиска в веб-интерфейсе АФ авторов-сотрудников.

Поиск с использованием АФ авторов-сотрудников работает по методу «от одного ко многим». Этот метод дает возможность связать некоторую группу терминов-синонимов, и при отборе на поиск одного из них, подменяет его в запросе на группу терминов, объединенных логикой «ИЛИ».

Первым этапом решения этой задачи стало наполнение имеющейся стандартной базы данных АФ индивидуальных авторов. Для ускорения этого процесса мы использовали конвертацию списков сотрудников, полученных из отдела кадров. Таким образом были созданы первоначальные записи с основным заголовком

– официальной формой имени лица на украинском языке. Следующим этапом было наполнение каждой записи ссылочными формами на других языках и разночтениями форм имени лица. Этот этап был связан с поиском данных об авторе по имеющимся записям в ЭК и работой с кафедрами, которые помогали нам, подавая списки своих сотрудников с переводами их имен на разные языки.

Далее начался этап создания самого поискового механизма, использующего в своей работе АФ. В решении этой задачи во многом помогла имеющаяся в веб-интерфейсе АБИС система предварительной проверки запросов. Эта сугубо служебная система обрабатывает все выполняемые запросы, проверяя и корректируя при необходимости в них служебные данные. Это так называемая система «Формально-логического контроля запросов» (CGIFLC). Фактически она дает возможность преобразовать весь запрос (со служебной и поисковой частью) до этапа его выполнения на сервере. Именно это ее свойство и было использовано в дальнейшей разработке нового поискового механизма.

Как уже говорилось ранее, суть поиска по АФ заключается в том, что один поисковый термин подменяется группой терминов. Однако далеко не всегда бывает необходимость получения всего списка работ автора. В отдельных случаях требуется получить результат контекстного поиска именно по той форме лица, которая указана в поисковом запросе. По этому, кроме реализации механизмов поиска требовалось реализовать возможность корректного его отключения. И тут на помощь пришло использование в веб-интерфейсе механизмов усечения, которые свидетельствуют о том, что запрос включает в себя не сам термин, а приближенную к нему форму. При

установке соответствующего признака, в запрос попадает определенный символ, который и запускает механизм поиска по АФ.

К сожалению, используя стандартный инструментарий и набор функций, предоставленный разработчиками АБИС Ирбис, реализовать задачу подмены терминов не получилось, и пришлось разрабатывать свой собственный модуль, главной задачей которого является разбиение поискового запроса на составляющие с возможностью определения поисковых словарей, в которых ищутся те или иные термины. С помощью доработанного функционала удалось выделить из поискового запроса всех авторов, обработать их определенным образом и поместить назад в поисковый запрос в преобразованном виде. Непосредственно преобразование имени лица в набор вариантов имен лиц осуществляется штатными средствами АБИС Ирбис – запросами и форматами. Значительным моментом оказалась возможность при анализе поискового запроса обрабатывать ряд введенных авторов, подбирая для каждого из них нужные формы лица.

На пути разработки этого механизма пришлось внести ряд изменений системы АФ в АБИС Ирбис. Так существенным изменением стало то, что АФ предполагает наличие отдельной записи на каждую принятую форму лица с перечнем прочих форм в списке ссылочных данных. Таким образом, на одного автора создавался целый ряд записей, что в конечном итоге затрудняло работу по поддержанию АФ в актуальном состоянии. Было принято решение отказаться от группы записей и оставить одну запись с перечнем всех форм. Таким образом, в словарях АФ остались только основные заголовки на украинском языке. Потребовался

ряд изменений для того, что бы обеспечить поиск нужной записи по любой из форм его имени.

Не обошлось и без ошибок на пути создания этого механизма. Так ошибки в алгоритме обработки поискового запроса без предварительной проверки привело к значительным затратам вычислительных ресурсов и частым отказам системы из-за длительности ответов. Но опытная эксплуатация и постоянный контроль результатов поиска понемногу дали возможность устранить эти проблемы.

Анализируя поисковые запросы, поступающие от веб-интерфейса АБИС Ирбис, за год эксплуатации читателями, было замечено, что количество результативных запросов значительно возросло. Часть возросшей результативности можно списать на то, что читатели не сразу смогли адаптироваться и перейти на новый поисковый интерфейс, что хорошо видно по ранним запросам, составленным с явными ошибками. Но так же значительный вклад в повышении результативности поиска внесло использование механизмов поиска по АФ.

На данный момент механизм использование АФ авторов-сотрудников работает в полной мере и приносит свои плоды, что видно при анализе результативности запросов. Однако останавливаться на достигнутом мы не собираемся, и в планах стоит доработка этого механизма до универсальности, позволяющей работать с любым поисковым словарем. Приоритетным на данный момент является направление работы с тезаурусом. Это обусловлено тем, что крайне малок количество читателей пользуются классификационными системами для поиска. Чаще всего при построении запросов ими используются сугубо ключевые слова, причем не в нормированной форме, а в той, в которой это слово встречается в рассматриваемой ими теме. Часто это

приводит к тому, что в результате поиска не возвращается вообще результаты, или они оказываются не полными. Ране в библиотеке было принято указывать на каждый отдельный документ как можно больше ключевых слов, описывающих его с разных сторон. В некоторой степени это компенсировало проблему неправильных или неполных запросов, но приводило к избыточному наполнению записей и лишней работе каталогизаторов. Теперь же при наличии механизма поиска по АФ, можно отказаться от избыточности. Результативность может быть достигнута за счет создания качественного тезауруса с раскрытым поисковым образом имеющихся в нем терминов. Это даст возможность правильно и полно подбирать документы по усеченному или ошибочному запросу.

В завершении хотелось бы сказать, что век информатизации, компьютеризации и использования мировых поисковых сервисов диктует свои условия и правила. Давно обленившийся читатель не хочет утруждать себя тщательным анализом искомой информации и изучением правил построения запросов. Он хочет быстро, многогранно и актуально получить всю информацию по своим весьма примитивным запросам. И если поисковые системы в Интернет давно решили эту проблему в свою пользу, то в АБИС это еще не пройденный этап, на пороге которого мы сейчас стоим.

***Аннотация:** в данной статье затрагивается тема средство унификации информации, используемой для каталогизации и поиска.*

***Анотація:** у даній статті зачінається тема засіб уніфікації інформації, використовуваної для каталогізації і пошуку.*