

# ИНФОРМАЦИОННО-ПОИСКОВЫЕ ЯЗЫКИ И СИСТЕМЫ

УДК 025.49

Г. А. Скарук

## Принципы комплексного использования лингвистических средств тематического поиска в электронном каталоге

*По итогам исследований в ГПНТБ СО РАН выведены основные принципы и правила комплексного использования в электронном каталоге нескольких информационно-поисковых языков разного типа.*

**Ключевые слова:** электронный каталог, лингвистическое обеспечение, информационно-поисковые языки, комплексное использование лингвистических средств, принципы поиска.

Использование разнообразных по своим характеристикам языков тематического поиска – одна из основных тенденций развития лингвистического обеспечения (ЛО) электронных каталогов. Однако восприятие совокупности информационно-поисковых языков (ИПЯ) как простого набора, где каждый язык применяется в отдельности, независимо от других лингвистических средств, существенно ограничивает возможности ЛО.

В ГПНТБ СО РАН с 1990-х гг. и по сей день проводятся исследования функциональных возможностей различных ИПЯ и поведения пользователей при поиске с их применением (например [1, 2]). Эти исследования позволили нам прийти к выводу: в ЭК должен функционировать именно комплекс поисковых языков. Свойства и принципы использования такого комплекса мы попытаемся рассмотреть в этой статье.

Комплекс (от лат. *complexus* – [связь, сочетание](#)) – сочетание, совокупность объектов, связанных между собой, образующих единую систему.

Основное требование к комплексу ИПЯ ещё на начальных этапах становления концепции ЛО АБИС формулировалось как *целостность*, заключающаяся в максимально возможном количестве вариантов взаимодействия и связанности всех ИПЯ и сохранении такой связанности при их использовании [3]. Помимо этого комплекс характеризуют такие свойства, как структурность, функционирование (как способ взаимодействия с другими системами и средой). С этих позиций ЛО представляется как совокупность объектов, взаимодействие которых обуславливает появление новых интегративных качеств, не свойственных образующим её компонентам.

Из этих положений вытекает *принцип взаимодополнения*. Поисковые возможности входящих в ЛО ЭК ИПЯ должны не повторять, а дополнять друг друга. Различные языки обнаруживают разную эффективность при реализации различных типов запросов. Часто недостатки ИПЯ являются оборотной стороной их достоинств. Комплексное использование лингвистических средств позволяет полнее реализовать преимущества и компенсировать недостатки языков.

Принцип взаимодополнения ИПЯ необходимо соблюдать: при определении состава и параметров ЛО ЭК в целом; при составлении поискового образа каждого отдельного документа (при этом необходимо учитывать, что при поиске в ЭК возможно не только комплексное, но и самостоятельное применение отдельных языковых средств, т.е. пользователь каталога может искать только на одном языке, например предметных рубрик, или использовать несколько ИПЯ, например, классификационный ИПЯ в сочетании с ключевыми словами); при составлении поискового образа каждого отдельного запроса.

В процессе поиска важно иметь в виду два основных принципа – дифференциации функций и интеграции функций ИПЯ.

*Принцип дифференциации функций отдельных ИПЯ.* При поиске лингвистические средства необходимо выбирать в зависимости от их функциональных свойств и возможностей, необходимых и достаточных в конкретной поисковой ситуации. Проведённое эмпирическое исследование продемонстрировало конкретные возможности применения этого принципа. В частности, выявлена целесообразность:

использования таблиц иерархических классификаций для уточнения запроса;

применения полного комплекса ИПЯ для реализации поиска с научными целями;

использования ключевых слов из заглавий с последующим обращением к авторитетным файлам контролируемых предкоординатных ИПЯ (иерархических классификаций, языка предметных рубрик) при поиске с учебными целями;

поиска посредством иерархической классификации для обеспечения наибольшей полноты выдачи, а также по запросам социально-гуманитарной тематики и содержащим сложные/множественные понятия;

поиска на языке предметных рубрик с целью обеспечения наибольшей точности, а также по запросам в области техники.

*Принцип интеграции функций ИПЯ.* В ряде случаев комбинация нескольких языковых средств при поиске даёт принципиально более качественные результаты, которые невозможно получить на основе любого из отдельных языков.

Например, для того чтобы выразить запрос «ультразвук в медицине» на вербальном поисковом языке (предметных рубрик, дескрипторов, ключевых слов), нужно перечислить все возможные сферы применения ультразвука и все названия заболеваний, для лечения которых он используется. Сделать это с исчерпывающей полнотой очень трудно. Поэтому сочетание термина *ультразвук* с наиболее широким классификационным индексом окажется более удобным. Правостороннее усечение кода ГРНТИ позволит получить в результате поиска все документы по медицине, в поисковый образ которых включено слово «ультразвук».

«ультразвук\$» \* «76\$»

ключевые слова из код ГРНТИ

заглавий и предметных рубрик «Медицина»

рубрик

С целью реализовать вышеназванные принципы поисковая ситуация описывается на основании *принципа комплексной диагностики*, согласно которому процедура диагностирования поисковой ситуации должна включать диагностику как самого запроса, так и предъявившего его пользователя. В обоих случаях необходима оценка комплекса характеристик: при диагностировании запроса – его отраслевой принадлежности, состава, содержания и объёма включённых в него семантических категорий, вида искомых документов; при диагностировании пользователя – цели поиска, уровня знания темы поиска, наличия поискового опыта.

Анализ состояния этих компонентов может помочь в уточнении качества запроса,

составленного пользователем, его требований к результатам поиска, выборе поисковых инструментов.

*Принцип комплексной оценки* свойств и функциональных возможностей отдельных ИПЯ применяется как на уровне конкретного поискового предписания, так и на уровне ЛО в целом. Рассматриваются возможности ИПЯ по обеспечению и эффективности, и комфортности [4] поиска. С этих позиций оценке подлежат:

*свойства самого ИПЯ*, в том числе: семантическая сила ИПЯ (состав и специфичность лексических единиц (ЛЕ), состав и функции грамматических средств, наличие механизмов многоаспектного представления содержания документа/запроса средствами ИПЯ); способность обеспечить результаты требуемого уровня качества для различных типов поисковых запросов;

*состояние и содержание методик индексирования* на различных ИПЯ, в том числе: степень адаптации к ситуации автоматизированного поиска, степень ориентации на потребности пользователя ЭК, предусмотренные ограничения в полноте и точности описания содержания документа/запроса;

*возможности манипулирования ИПЯ неподготовленными пользователями ЭК*, в том числе: объём и структурная сложность лексики и грамматики поискового языка, наличие механизмов экспликации смысла ЛЕ, наличие механизмов экспликации структуры и состава ИПЯ, степень соответствия структуры ЛЕ и нормативных словарей ИПЯ естественным речемыслительным структурам, наличие механизмов уточнения темы запроса;

*состав и содержание типичных ошибок при индексировании документов и запросов*;

*возможности представления и коррекции свойств ИПЯ программными средствами*, в том числе: возможности поиска с усечением, использования в режимах пред- и посткоординации, доступа к авторитетным файлам ИПЯ.

Как составляющую принципа дифференцированного использования можно рассматривать *принцип дифференцированного доступа*, в соответствии с которым степень раскрытия для пользователей состава и структуры ИПЯ должна определяться по результатам диагностики поисковой ситуации. В зависимости от квалификации пользователя каталога, уровня сложности и категории запроса одним читателям будет необходим просмотр нормативных словарей ИПЯ (авторитетного файла предметных рубрик, электронных или печатных таблиц классификации), для других будет достаточно непосредственно ввести текст запроса в поисковое предписание.

Как показал эксперимент, для успешной реализации дифференцированного подхода к ИПЯ необходимо *сопряжение отдельных ИПЯ* в составе комплекса. В электронном каталоге это означает установление смыслового соответствия между лексическими единицами разных ИПЯ. Более того, должна быть предоставлена возможность автоматического перехода с одного языка поиска на другой. Например, в ГПНТБ СО РАН сопряжение языка предметных рубрик и ББК на первом этапе решено осуществить с помощью Авторитетного файла предметных рубрик [5]. Каждая авторитетная запись помимо предметной рубрики будет содержать соответствующий индекс ББК.

Из принципа комплексной оценки закономерно вытекает *принцип комплексного преобразования ЛО ЭК*, согласно которому в случае преобразования одного из компонентов ЛО следует оценивать его влияние на все перечисленные при формулировании принципа

комплексной оценки характеристики каждого из элементов ЛО и при необходимости осуществлять согласованное преобразование всего комплекса ИПЯ.

Оценку результатов применения названных принципов целесообразно проводить на основе *принципа комплексного контроля качества индексирования*. Анализ ошибок, допущенных при индексировании документов, позволяет говорить о необходимости в процессе контроля качества результатов семантической обработки соотносить состав и структуру поискового образа документа (ПОД) на различных ИПЯ. Эталоном в этом случае выступает ПОД, в результате декодирования которого получено наиболее полное описание содержания документа.

Важно оценить: соотношение смыслов и объёмов понятий, полученных в результате декодирования, соотношение количества ЛЕ, использованных для описания основного предмета и аспектов содержания документа. Далее оценивается, имеется ли возможность отразить выявленные пробелы в описании содержания средствами каждого поискового языка и целесообразно ли такое дополнение в ситуации комплексного использования ИПЯ. Правила взаимодополнения индексов на различных ИПЯ должны быть чётко регламентированы в соответствующих нормативно-методических документах и учтены при составлении ПОД.

Из последнего положения следует *принцип координации методик индексирования отдельных ИПЯ* с учётом как специфических функций ИПЯ в составе комплекса, так и задач их сопряжения.

По нашему мнению, соблюдение описанных принципов будет способствовать эффективному и экономичному использованию комплекса ИПЯ в целом и отдельных поисковых языков в его составе.

### **Список источников**

1. **Скарук Г. А.** Комплексное использование лингвистических средств тематического поиска в электронном каталоге / Г. А. Скарук // Библиотекосведение. – 2010. – № 2. – С. 34–40.
2. **Скарук Г. А.** Пользователь у каталога: результаты социологического исследования / Г. А. Скарук, Л. А. Жарикова, Т. В. Бусыгина // Науч. и техн. б-ки. – 2009. – № 6. – С. 5–18.
3. **Гендина Н. И.** Лингвистическое обеспечение автоматизированных библиотечных систем. – Алма-Ата, 1991. – 222 с.
4. **Скарук Г. А.** Роль лингвистических средств в обеспечении комфортности поиска в электронном каталоге / Г. А. Скарук // Науч. и техн. б-ки. – 2011. – № 1. – С. 101–110.
5. **Скарук Г. А.** Авторитетный файл предметных рубрик: новые возможности индексирования и поиска [Электронный ресурс] / Г. А. Скарук // Научные библиотеки России: взгляд в будущее : межрегион. науч.-практ. конф. (Новосибирск, 26–28 сентября 2012 г.). – Новосибирск, 2012.