

Г. А. Скарук
ГПНТБ СО РАН

Возможности поиска по классификационным индексам в электронном каталоге

Рассмотрены поисковые возможности классификационных ИПЯ в электронной среде, как унаследованные от карточных систематических каталогов, так и приобретённые в условиях автоматизированного поиска. Освещены проблемы, возникающие при поиске по классификационным индексам, и пути их решения.

Ключевые слова: классификационные информационно-поисковые языки, иерархические классификации, автоматизированный поиск, электронный каталог, поисковые возможности.

UDC 025.4.03

Galina Skaruk

*Russian Academy of Science Siberian Branch State Public Library
for Science and Technology, Novosibirsk, Russia*

Options for searching e-catalog by classification numbers

The potential of classification information retrieval languages in digital environment (both the functions inherited from card systematic catalogs and the new ones acquired for computerized environment) is examined. The problems of search by classification numbers are addressed and the solutions are offered.

Keywords: classification information retrieval languages, automated search, electronic catalog, search functionality.

В российских библиотечных каталогах классификационные системы долгое время были основным средством поиска и вполне справлялись со своими задачами. Отечественные специалисты накопили большой опыт индексирования и поиска посредством иерархических классификаций. Классификационные индексы (чаще всего ББК и УДК) и сейчас присутствуют в библиографических записях большинства электронных каталогов, однако

по многим причинам возможности поиска с их использованием в существующих сегодня ЭК оказались очень ограничены по сравнению с карточными систематическими каталогами.

Среди таких причин – дефицит программных разработок высокого уровня для представления и использования классификаций в электронной среде, предполагающих вербальный доступ к классификационным индексам; трудоёмкость и ресурсоёмкость подобных разработок; рассмотрение классификационных ИПЯ как «наследия социализма». Не на последнем месте и широко распространившееся сегодня представление о простоте, элементарности процессов информационного поиска и, на наш взгляд, нехватка у библиотечных специалистов информации о возможностях классификационного поиска в электронной среде.

В результате в ЭК мы не только утрачиваем поисковые возможности, которые нам предоставлял систематический каталог, но и теряем многие уникальные функции поиска, реализацию которых могло бы обеспечить грамотное применение классификационных индексов в автоматизированном режиме.

Далее попытаемся обозначить эти «старые» и новые возможности и проиллюстрировать их конкретными примерами.

Поиск по широким темам

Поиск по широким темам всегда был проблемой при использовании вербальных (словарных) языков в автоматизированных поисковых системах. Если пользователь, обратившись к каталогу, хотел получить не только «общие работы» по теме в целом, но и «всё, что есть по этой теме в библиотеке», то он был вынужден указывать в поисковом предписании (ПП) все предметные рубрики или ключевые слова, которые, по его предположению, могут оказаться подчинёнными по смыслу термину, приведённому в наименовании общей темы. Например, если нужна полная информация о разработке месторождений всех цветных металлов, то для получения полной выдачи недостаточно привести рубрику *Цветные металлы – Добыча*. В ПП необходимо учесть, во-первых, наименования руд всех цветных металлов (медные, никелевые, свинцовые, цинковые и т.д.), во-вторых, все возможные аспекты рассмотрения темы в разрезе горного дела (все системы и методы разработки месторождений, оборудование для разработки и т.д.).

Приведу лишь фрагмент такого ПП: *Руды цветных металлов – Добыча + Медные руды – Добыча + Никелевые руды – Добыча + Кобальтовые руды – Добыча + Свинцовые руды – Добыча + Цинковые руды – Добыча + Оловянные руды – Добыча, Ртутные руды – Добыча + Ртутные руды – Добыча – Висмутовые руды – Добыча + Аллюминиевые руды – Добы-*

ча + Бокситовые руды – Добыча + Медные руды – Месторождения – Системы разработки с закладкой выработанного пространства + Медные руды – Добыча – Буровые работы + Медные руды – Добыча – Открытые работы.

Отметим: полное ПП могло бы занять несколько страниц.

Классификационный индекс с тем же значением (например, индекс ББК И333¹) «работает» в автоматизированном поиске совершенно по-другому. С помощью классификационного ИПЯ мы имеем возможность в кратком ПП объединить большое число понятий по отраслевой теме. Процедура поиска предполагает автоматическое включение в выдачу всех лексических единиц, иерархически подчинённых приведённым в запросе. Поэтому, если мы зададим ПП *И333*\$, в выдачу по запросу войдут документы не только с индексом *И333*, но и со всеми индексами, нижестоящими по отношению к данному, и, соответственно, темами:

И333 Разработка месторождений руд цветных металлов

И333.2 Разработка месторождений руд тяжелых цветных металлов

И333.21 Разработка месторождений медных руд

И333.22 Разработка месторождений никелевых и кобальтовых руд

И333.24 Разработка месторождений свинцовых и цинковых руд

И333.26 Разработка месторождений оловянных руд

И.333.27 Разработка месторождений ртутных, сурьмяных и висмутовых руд

И 333.3 Разработка месторождений руд легких металлов

И333.4 Разработка месторождений руд благородных металлов и т.д.

Оба поисковых предписания – *И333* и *Руды цветных металлов – Добыча* – выражают один и тот же смысл и объём понятия, но при поиске с усечением обеспечивают совершенно разные результаты. При этом, меняя длину индекса путём усечения в ту или иную сторону, мы либо сужаем, уточняем, либо расширяем тематические рамки обзора.

Такой поиск даёт возможность, не изменяя объём отражённого в индексе понятия, существенно расширить поле обзора и объём выдачи. А значит, в иерархических классификациях изначально решена типичная для поиска на вербальных ИПЯ «проблема общих и частных терминов». Подобными возможностями в таком объёме не располагает ни один словарный поисковый язык, используемый в ЭК.

¹ Здесь и далее в примерах используются индексы Полных таблиц ББК в 25 томах, на которых базируются систематический и электронный каталоги ГПНТБ СО РАН.

Поиск по запросам, содержащим категории пространства и времени

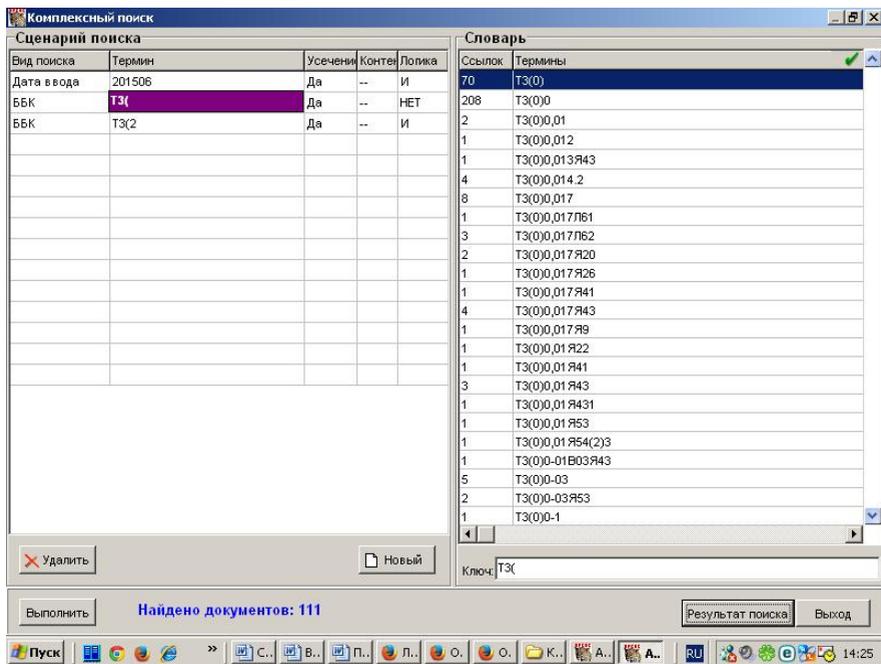
Поиск по этим тематическим категориям всегда представлял собой особую проблему. Над тем, как отразить, а потом найти материал обо всех исторических событиях в рамках широкого исторического периода или обо всех населённых пунктах в рамках обширной территории, библиотечные специалисты задумывались не один раз.

Так, например, в методических материалах по использованию хронологических заголовков в предметных рубриках [2] Л. В. Завьялова подчёркивает особую сложность вопроса об адекватности данного типа подзаголовков, призывает избегать предметизации документов с использованием точных дат в тех случаях, когда документов с этой датой в каталоге немного, так как пользователям трудно учесть все даты при поиске по более широким историческим периодам.

По нашему мнению, руководствуясь этим правилом, нельзя забывать о том, что ЭК содержит описания документов за многие годы, и достаточно трудно предсказать, как через несколько лет изменится количество изданий по той или иной теме. Поэтому ориентация на наполнение раздела каталога на данный момент времени при определении степени адекватности описания хронологического периода представляется достаточно спорным решением. Выходом в этой ситуации может стать использование классификационных индексов при поиске по широким историческим периодам. Приведу пример.

Читателю необходимы книги по истории российского казачества в XVII-XVIII вв. Тему адекватно отражает индекс Полных таблиц ББК "ТЗ(2)46-294" (где 46 – хронологическое типовое деление, означающее века), но документы с этим индексом в ЭК не найдены. Используя вышестоящий индекс "ТЗ(2)4-294" (где хронологическое деление 4 отражает период с VIII в. по 1861 г.) читатель получил две книги, содержащие и информацию об указанном в запросе историческом периоде.

То же можно сказать и о роли классификационных ИПЯ при поиске, в котором необходимо охватить все документы о достаточно широкой территории. Отметим, что с помощью классификации в этом случае можно реализовать запросы, которые при поиске с использованием других лингвистических средств выполнить чрезвычайно трудно. Например, читателю нужна литература по истории всех зарубежных стран, поступившая в библиотеку за последний месяц. Практически невозможно включить в ПП названия всех зарубежных стран, не говоря уже о более мелких населённых пунктах. Решить проблему можно, сформулировав ПП на языке, например, ББК.



В таком случае в выдачу будут включены описания документов по истории всех стран, кроме России.

Поиск по отдельным элементам индекса

В электронной среде появляется возможность использовать не только классификационные индексы как единую совокупность знаков, максимально отражающих содержание документа, но и иерархическую классификацию как посткоординатную систему, что подразумевает использование всех фрагментов полного индекса как независимых поисковых элементов. Все элементы можно сочетать в поиске с использованием булевых и контекстных операторов. К каждому элементу могут быть применены операции право- и левостороннего усечения. Это означает, что при усечении в сферу поиска будут включены подчинённые приведённым в ПП индексы как основных, так и вспомогательных таблиц.

Иллюстрируя это положение, продолжим пример поиска по истории казачества. В случае поиска по отдельным элементам индекса (основных таблиц, таблиц территориальных типовых делений, таблиц специальных типовых делений) можно построить ПП, используя усечение: «ТЗ(2\$) * «\$–

294\$». При такой формулировке в выдачу войдут документы не только о всех более широких исторических периодах, но и об истории казачества различных регионов и населённых пунктов России.

Возможности поэлементного поиска по индексам комбинационных классификаций были описаны ещё в 1976 г. [4–7], однако в отечественных ЭК эта стратегия до сих пор не реализована на практике.

Ограничение области поиска

Классификационные ИПЯ очень полезны при поиске по узкой теме в широком отраслевом контексте.

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 ключевого слова (в ЭК ГПНТБ СО РАН это любое слово из любого поля библиографической записи) «ультразвук» с широким классификационным индексом окажется наиболее удобным и экономичным вариантом.

Проще всего такой широкий контекст задать с помощью Государственного Рубрикатора научно-технической информации (ГРНТИ). 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 «Медицина»
рубрик,*

где «*» – знак логического произведения «И».

«Карты знаний». Одна из функций, контролируемых ИПЯ в ЭК, – визуализация содержания библиотечных фондов. Задача поиска и просмотра может быть облегчена с помощью использования заранее созданного комплексного механизма (карта знаний, онтология понятий, сеть знаний) [10]. Поэтому всё чаще предлагается преобразовать структуру традиционных поисковых языков в соответствии с новыми требованиями. Так, даже списки предметных рубрик рекомендуется представлять в виде древовидных графов [8, 9].

Напомним, что в классификационных ИПЯ иерархическая древовидная структура заложена изначально. Именно иерархические классификации издавна рассматривались как «карты знаний» [1, 6], способные целостно и наглядно представить объекты, обозначить связи и размежевания между

ними. Это значительно облегчает просмотр словарей ИПЯ и отбор терминов для составления запроса.

Названный способ экспликации структуры и лексики ИПЯ кажется нам очень перспективным с учётом того, что иерархия обоснованно считается одной из наиболее простых и удобных форм представления знаний, соответствующих естественным механизмам речемыслительной деятельности человека.

Структурирование выдачи. При поиске в базах данных большого объёма, каковыми сегодня являются многие электронные каталоги, достаточно часто (особенно, если задать нечёткое ПП) выдача достигает нескольких тысяч документов. В подобных случаях очень полезна такая функция классификационных ИПЯ, как тематическое структурирование выдачи. Это помогло бы читателям точнее очертить область интереса и облегчило бы просмотр.

Кроме того, такая организация выдачи позволит читателям изучить не только обозначенные ими темы, но и их семантическое окружение, увидеть новые аспекты рассмотрения, обнаружить «пробелы» в знании темы, которые очень трудно выразить словами. То есть, читатель реализует потребность «рыться» в книгах, творчески осмысливать свою тему в процессе библиографического поиска, о чём не раз говорил Э. Р. Сукиасян [6]. Однако такая возможность пока остаётся гипотетической.

Языки-посредники. Нотация классификационных систем не зависит от конкретной системы письма на том или ином естественном языке. Это свойство позволяет классификационным языкам выполнять в современной сетевой электронной среде роль языков-посредников. Во-первых, условные семантические коды, применяемые в классификационных системах, могут связывать между собой различные вербальные ИПЯ. Во-вторых, международные классификационные системы могут использоваться для создания единого входа в распределённых многоязычных системах поиска, объединяющих ЭК и другие электронные ресурсы многих стран.

Проблемы поиска с использованием классификационных ИПЯ. Безусловно, при поиске с использованием классификационных ИПЯ возникают проблемы, обусловленные особенностями их структуры и нотации. Однако, говоря об естественных ограничениях поиска, нельзя упускать из вида два момента:

1) многие недостатки классификаций как поисковых языков в электронной среде могут быть преодолены;

2) ряд ограничений классификационных ИПЯ является «оборотной стороной» их преимуществ, что в условиях комплексного использования ИПЯ [5] компенсируется за счёт применения лингвистических средств других типов.

Попытаемся кратко обрисовать основные проблемы и предлагаемые сегодня способы их преодоления.

Трудность понимания пользователями содержания классификационных индексов чаще всего указывается как причина недостаточного использования классификаций в ЭК. В связи с этим уже давно обсуждается создание «предметного входа» в классификационные таблицы, т.е. устранение названной причины. Возможны два варианта организации такого входа:

обеспечение вербального доступа к классификационным делениям непосредственно в «электронных таблицах» или авторитетных файлах классификационных данных;

сопряжение классификационных языков с вербальными ИПЯ.

Первый вариант можно увидеть в «навигаторах» ББК, УДК, ГРНТИ, разработанных ГПНТБ России (http://librarynew.gpntb.ru/cgi/irbis64r_simplesite/cgiirbis_64.exe?C21COM=F&I21DBN=IBIS&P21DBN=IBIS&S21CNR=&Z21ID=).

Пример второго подхода – *Classification Web* (электронные таблицы Классификации Библиотеки Конгресса США – КБК; <https://classificationweb.net/>). В этой системе установлено соответствие между индексами двух классификаций – КБК и Десятичной классификации Дьюи – и предметными рубриками Библиотеки Конгресса США (LCSH) [3].

О проблеме **недостаточной гибкости классификационных ИПЯ** обычно говорят, сравнивая их с посткоординатными языками. Речь идёт о невозможности в рамках жёсткой классификационной системы проводить поиск по произвольному сочетанию поисковых признаков. Однако описанная нами выше возможность поиска по отдельным элементам классификационных индексов позволяет преодолеть и этот недостаток, порождённый линейным представлением индексов. И всё же при поиске по отдельным элементам возникает другая проблема, которую рассмотрим далее.

Информационный шум при поиске по элементам классификационных индексов. Мы знаем, что, в зависимости от места в таблицах одни и те же цифровые (буквенно-цифровые) сочетания могут иметь совершенно разное значение. Например, в таблицах ББК по техническим наукам СТД -7 означает технический контроль, а в клинических разделах медицины – «осложнение отдельных заболеваний и их лечение». Поэтому при поиске по

индексам специальных типовых делений (СТД) в ББК или специальных определителей в УДК может возникнуть чрезмерный поисковый шум. В этом случае выход – стандартизация и унификация системы типовых делений, их межотраслевое согласование.

Ограничения поиска, обусловленные методикой систематизации.

В своё время методики систематизации посредством различных ИПЯ разрабатывались в расчёте на карточные каталоги. Многие методические приёмы тогда формировались с учётом особенностей расположения карточек, поиска и просмотра, принятых в карточных систематических каталогах. Поэтому практически во всех методиках систематизации мы можем найти правила, позволяющие сделать полный индекс более компактным.

К подобным правилам можно отнести обобщение при определённых условиях нескольких тем под более широким индексом, «скрытые» индексы (в УДК и ГРНТИ) – последовательности индексов, обозначенные через знак распространения, определение степени детализации индексов в зависимости от наполнения разделов каталога. В условиях последовательного поиска в каталогах «карточка за карточкой» эти приёмы были допустимы и оправданы. Путём последовательного просмотра пользователь находил в общих разделах каталогов описания книг по более конкретным темам.

Но в электронных каталогах поиск происходит по другим принципам. В результате поиска читателю выдаются описания документов только с конкретным индексом, который был приведён в ПП. Это означает, что при систематизации для ЭК необходимо повысить полноту индексирования, стремиться отразить в индексе все рассматриваемые в документе темы. Например, в УДК при систематизации для автоматизированного поиска не следует использовать символ распространения. Вместо 546.32/.35

Наиболее важные галогены следует перечислить все составляющие:

546.32 *Калий*

546.33 *Натрий*

546.34 *Литий*

546.35 *Рубидий*

Очевидно, для того чтобы сделать классификационные индексы эффективным инструментом автоматизированного поиска, необходим достаточно серьёзный пересмотр методики систематизации.

Подведём итоги:

1. Благодаря особенностям нотации и структуры, классификационные ИПЯ в электронной среде приобретают уникальные поисковые возможности, которые не в состоянии обеспечить ни один вербальный поисковый язык.

Поэтому исключение их из числа инструментов поиска в электронных каталогах приведёт к серьёзным функциональным ограничениям ЭК.

2. Функционирование классификационных индексов в ЭК принципиально отличается от поиска на вербальных ИПЯ. Самые серьёзные изменения произошли в связи с появлением возможностей поиска по отдельным элементам индексов, усечения индексов и использования логических операторов.

3. Наиболее эффективным представляется функционирование классификационных систем в составе комплекса лингвистических средств ЭК. Встроенные в систему комплексного поиска классификационные языки смогут компенсировать недостатки словарных ИПЯ.

4. Большинство свойств классификационных ИПЯ, снижающих эффективность их использования, могут быть изменены в электронной среде. Однако эти преобразования требуют значительных затрат, как ресурсных, так и интеллектуальных, поэтому они могут быть проведены только путём консолидации усилий библиотечного сообщества.

5. Для эффективного использования в ЭК необходима модификация лексического состава и грамматики классификационных ИПЯ. Основные направления преобразований: унификация структуры и лексики классификационных таблиц, оптимизация структуры систем классификации с использованием принципов фасетизации, формализация использования грамматических средств с целью обеспечения однозначности их интерпретации в ЭК, существенное расширение ссылочно-справочного аппарата.

6. Ориентированная на карточные каталоги методика систематизации ограничивает поисковые возможности иерархических классификаций в электронной среде. Необходима её адаптация с целью повышения полноты и точности поиска посредством классификационных ИПЯ.

СПИСОК ИСТОЧНИКОВ

1. Гендина Н. И. ББК как средство организации знаний и предмет изучения в высшей библиотечной школе: концептуальные подходы // ББК: Новые горизонты организации знаний : материалы Международ. науч. конф. (14–16 окт. 1998 г., Москва). – Москва, 1999. – С. 86–100.

Gendina N. I. BBK kak sredstvo organizatsii znaniy i predmet izucheniya v vysshey bibliotечноy shkole: kontseptualnye podhody // BBK: Novye gorizonty organizatsii znaniy : materialy Mezhdunarod. nauch. konf. (14–16 okt. 1998 g., Moskva). – Moskva, 1999. – S. 86–100.

2. Завьялова Л. В. Хронологические подзаголовки. – Режим доступа: http://www.nlr.ru:8101/cat/preddmet/method/obmet/subtitle_chron.pdf (Дата обращения 27.07.15).

Zavyalova L. V. Hronologicheskie podzagolovki. – Rezhim dostupa: http://www.nlr.ru:8101/cat/predmet/method/obmet/subtitle_chron.pdf (Data obrashcheniya 27.07.15).

3. **Каталогизация.** Современные технологии. Тенденции и перспективы развития / Ю. Г. Селиванова [и др.]; Нац. информ.-библ. центр ЛИБНЕТ, Рос. нац. б-ка, Рос. библ. ассоц. – Москва : ФАИР : Центр ЛИБНЕТ, 2007. – 215 с.

Katalogizatsiya. Sovremennye tehnologii. Tendentsii i perspektivy razvitiya / Yu. G. Selivanova [i dr.]. ; Nats. inform.-bibl. tsentr LIBNET, Ros. nats. b-ka, Ros. bibl. assots. – Moskva : FAIR : Tsentr LIBNET, 2007. – 215 s.

4. **Мишин В. А.** Возможности машинного поиска информации по классификационным индексам / В. А. Мишин // Организация систематических и предметных каталогов научных библиотек : сб. науч. тр. / Гос. б-ка СССР им. В. И. Ленина. – Москва, 1976. – С. 49–55.

Mishin V. A. Vozmozhnosti mashinnogo poiska informatsii po klassifikatsionnym indeksam / V. A. Mishin // Organizatsiya sistematicheskikh i predmetnykh katalogov nauchnykh bibliotek : sb. nauch. tr. / Gos. b-ka SSSR im. V. I. Lenina. – Moskva, 1976. – S. 49–55.

5. **Скарук Г. А.** Принципы комплексного использования лингвистических средств тематического поиска в электронном каталоге / Г. А. Скарук // Науч. и техн. б-ки. – 2014. – № 2. – С. 72–76.

Skaruk G. A. Printsipy kompleksnogo ispolzovaniya lingvicheskikh sredstv tematicheskogo poiska v elektronnom kataloge / G. A. Skaruk // Nauch. i tehn. b-ki. – 2014. – № 2. – S. 72–76.

6. **Сукиасян Э. Р.** За что читатели любили систематический каталог? / Э. Р. Сукиасян // Там же. – № 5. – С. 45–51.

Sukiasyan E. R. Za chto chitateli lyubili sistematicheskii katalog? / E. R. Sukiasyan // Tam zhe. – № 5. – S. 45–51.

7. **Сукиасян Э. Р.** Классификационная система как инструмент автоматизированного поиска. Часть 2. Поэлементный классификационный поиск / Э. Р. Сукиасян // Там же. – 2015. – № 4. – С. 53–62.

Sukiasyan E. R. Klassifikatsionnaya sistema kak instrument avtomatizirovannogo poiska. Chast 2. Poelementnyu klassifikatsionnyu poisk / E. R. Sukiasyan // Tam zhe. – 2015. – № 4. – S. 53–62.

8. **Donahue T.** Animated subject maps for book collections / Tim Donahue // Inf. Tech. and Libr. – 2013. – Vol. 32, N 2. – P. 7–17.

9. **Julien C.-A.** Capitalizing on information organization and information visualization for a new-generation catalogue / Charles-Antoine Julien, Catherine Guastavino, France Bouthillier // Libr. Trends. – 2012. – Vol. 61, N 1. – P. 148–161.

10. **Li Y. M.** A recommender mechanism for social knowledge navigation in an online encyclopedia / Yung-Ming Li, Lien-Fa Lin, Yu-Hui Lin // Inf. Proc. and Manag. – 2014. – Vol. 50, N 5. – P. 634–652.

Galina Skaruk, *Cand. Sci. (pedagogu), researcher, State Public Library for Science and Technology of the Siberian Branch of the Russia Academy of Science, Novosibirsk, Russia;*

skaruk@spsl.nsc.ru

15 Voskhod str., Novosibirsk, 630200 Russia