ХАРКІВСЬКА ДЕРЖАВНА АКАДЕМІЯ КУЛЬТУРИ

КАФЕДРА ІНФОРМАЦІЙНО-ДОКУМЕНТНИХ СИСТЕМ

Л.Я. ФІЛІПОВА

ТЕОРЕТИЧНА ІНФОРМАТИКА

Конспекти лекцій до курсу

(Розділ 1. Теоретичні основи інформаційної діяльності)

[Електронний ресурс]

Харків, ХДАК, 2014

СОДЕРЖАНИЕ:

Тема 1: Информатика как наука. Концепции информатики.

1. Предпосылки возникновения информатики.
2. Определения информатики как науки.
3. Основные концепции информатики.
4. Взаимосвязи с другими науками.

Список литературы:

1. Информатика: Учеб.пособие / Под ред. К.В. Тараканова — М.: Книга, 1986.— С. 5-17.
2. Хохлова Н.В. и др. Информатика: Учеб. пособие для вузов. — Мн.: Выш. шк., 1990.—С.9-22.
3. Михайлов А.И., Черный А.И., Гиляревский Р.С. Научные комуникации и информатика / ВИНИТИ.— М.:Наука ,1976.— С.71-117; 392-416.
4. Каныгин Ю.М., Калитич Г.И. Основы теоретической информатики.— К.: Наук. думка 1990.— 232 с.
5. Колин К.К. Эволюция информатики и проблемы формирования нового комплекса наук об информации // НТИ., сер.1 — 1995.— №5.— С. 1—7.

1. Предпосылки возникновения информатики.

Появление информатики как науки в обществе обусловлено многими факторами, основные из которых связаны с потребностями общественной практики, потребностями компьютеризации различных сфер производства. Важнейшие предпосылки возникновения информатики обусловлены закономерностями развития науки и своими корнями уходят в глубокую древность, в античную цивилизацию. Именно в те времена, особо разительными были успехи античной науки: философии, физиологии, географии, зоологии, минералогии, фармакологии, физиологии, математики, инженерных изысканий и др.

Накопление научного знания в эпоху классической европейской и далее – арабской цивилизации изменило сложившееся веками представление о некоей единой науке: начался процесс дифференциации наук по предмету исследования. В эпоху Возрождения этот процесс ускорился и завершился к началу XX в.

В ходе дальнейшего развития науки как системы знаний проявились диалектически взаимосвязанные, но противоположные тенденции: дифференциация и интеграция наук. Складывается стройная система научных знаний. Акад. Б. Кедров отразил ***архитектонику наук*** с помощью так называемого «треугольника наук» (рис.1.)

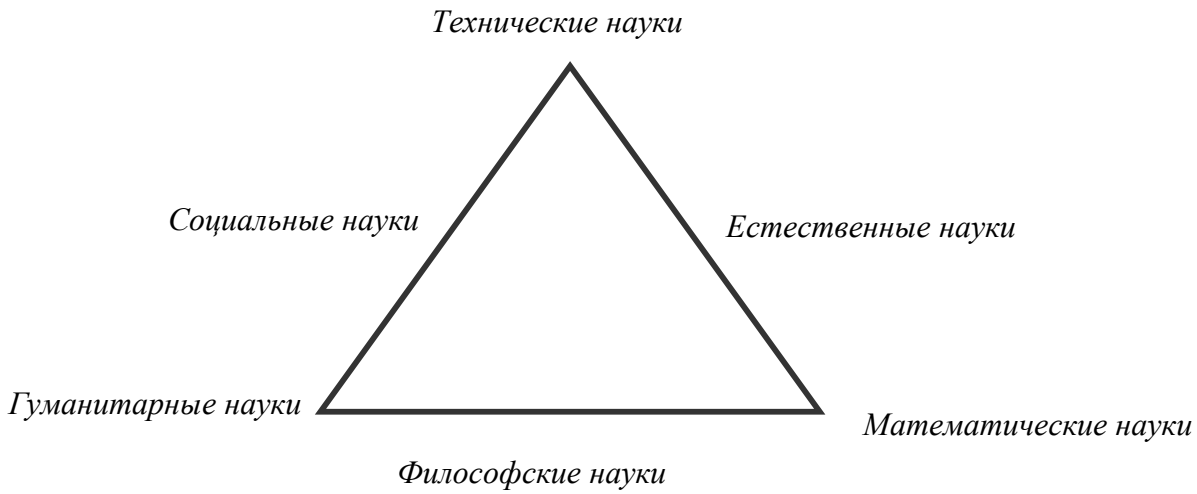

Рис.1. Архитектоника наук, выраженная «треугольником наук» (акад. Б. Кедров)

Три грани – это естественные, общественные (социальные) и философские науки, исследующие, соответственно, многообразные аспекты живой и неживой природы, общество (социум) и отрасли его материального производства, законы познания, восхождения от простого к сложному, от незнания к знанию. На стыке естественных и социальных наук возникли и развиваются технические науки (теория машин и механизмов, сопротивления материалов и др.), на стыке естественных и философских – математические (линейное и динамическое программирование, исследование операций, теория массового обслуживания и др.) на стыке социальных и философских наук – гуманитарные науки (языкознание, литературоведение и др.).

Таковы результаты общественного разделения наук по объекту и предмету исследования.

С середины прошлого столетия становится устойчивой тенденция функционального разделения труда в науке. Единству теории и практики как двух сторон научной деятельности соответствует разделение каждой науки на теоретическую и прикладную.

Через систему общественного разделения труда научная деятельность связана со всеми сферами социальной деятельности. Дифференциация наук по объекту и предмету исследования, а также по функциональному принципу привела к увеличению связей между науками. У ученых возникает объективная необходимость в обмене научными знаниями, в распространении результатов научного труда. Поэтому функция информирования обособилась, научно-информационная деятельность (НИД) стала специфической разновидностью научного труда, являясь его органической составляющей (рис 2.).

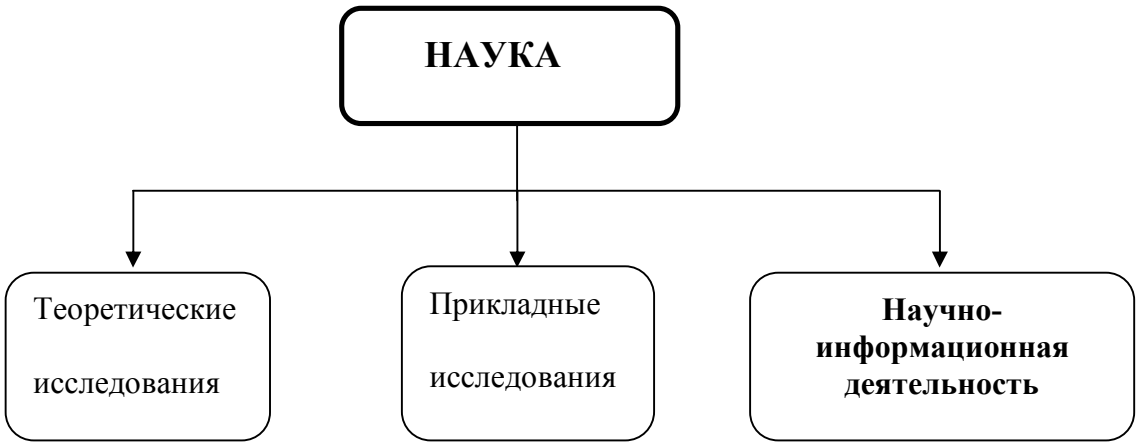

Рис.2. Общественное разделение труда в науке по функциональному признаку.

В то же время одной из отрицательных тенденций в науке стал информационный кризис сопровождаемый усложнением содержания научной информации и увеличением количества научных документов (опубликованных и не опубликованных), как материальных носителей научной информации, ассигнований на научные исследования, разработки и численности занятых ими. Это объективно привело к физической невозможности выявить и изучить необходимые источники на традиционных и принципиально новых материально новых носителях информации.

Следствием такого положения стали неоправданная трата времени, средств и сил. Эти негативные проявления наблюдались в увеличении затрат времени профессионалов на рутинные процессы поиска информации, параллелизм и дублирование научно-исследовательских работ (НИР) и опытно-конструкторских разработок (ОКР) с нежелательными экономическими потерями: поиском найденного, решением решенного, утратой новизны, патентно- и конкурентоспособности, патентной чистоты.

Информационный кризис кроме того, обострился возникновением информационных барьеров. Назовем некоторые из них:

1. Межъязыковые барьеры. В результате того, что к науке стали приобщаться все новые народы, в мировые научно-технической литературе увеличилось количество публикаций на таких языках, которыми не владеет большинство ученых. Так, если в начале XX в. на трех основных европейских языках: английском, немецком, французском, выходило 93,5 % всей научной литературы, то в последние годы она стала выходить на более 70 языках.
2. Терминологические барьеры. Представлены двумя разновидностями:

   а) внутриязыковой барьер. Вследствие увеличивающегося объема знаний и специализации наук происходит дифференциация научных языков, специализация терминологии, используемой в каждой из отраслей наук. Это приводит к тому, что

специалисты разных отраслей науки и техники все в меньшей степени понимают друг друга.

б) «Семиологический» барьер. В противовес искусственному дроблению единой науки, последняя стремится к объединению на более высоком уровне абстракции. По мере увеличения абстрактности в науке начинают применяться все более сложные и трудные для понимания понятия, специальная терминология. Вследствие этого число ученых, в полной мере понимающих научные теории, имеет тенденцию к относительному уменьшению. Эта же абстрактность науки порождает и еще одно отрицательное явление – потерю конкретного знания, имеющего для достижения практических целей наибольшую ценность.

3. Барьер чтения. Обусловлен ограниченностью физических возможностей человека. Человек читает со скоростью 200-300 слов в минуту при 60 % усвоении прочитанного материала. Подсчитано, что при ежедневном чтении 50 страниц за всю жизнь можно прочесть от 1 до 10 тыс. книг, тогда как за период человеческой жизни их выходит до 20 млн. экз. (данные 90-х гг. прошлого века).
4. Барьер доступности. Обусловлен недостатками в непосредственном процессе передачи информации либо между отправителем или потребителем, либо в системе информационного обслуживания, когда потенциальный потребитель, зная о существовании нужного документа, не имеет возможности им воспользоваться.
5. Коммуникативный барьер. Понимается как искажение и потеря информации при переработке ее в информационных службах.

К основным путям преодоления информационного кризиса относятся: усиление роли и повышение уровня профессионализма информационных работников как посредников между информацией и потребителем, развитие информационно- поисковых систем и др.

Появление информатики как науки, исследующей научно-информационную деятельность, явилось кульминацией и логическим завершением тех тенденций в науке, которые развивались в последние десятилетия XX в. Возникновение информатики обусловило переход к новым информационным технологиям, ставшим органичным элементом различных социально-комуникативных процессов – проектных разработок, кассовых операций и многих других.

Термин “информатика” как научная дисциплина имеет многозначное толкование, что вылилось в отдельные концепции информатики. Этот термин не имеет четкого

определения ни в одном нормативном документе. Нет единой точки зрения на определение этой науки как у отечественных, так и у зарубежных ученых.

2. Определения информатики как науки.

Приоритет первого определения в отечественной информатике принадлежит известным ученым Михайлову А.И., Черному А.И. и Гиляревскому Р.С. (ВИНИТИ): информатика — это “научная дисциплина, изучающая структуру и общие свойства информации, а также закономерности всех процессов научной коммуникации”. Ими же были внесены уточнения в определение информатики для социальной сферы — эта научная дисциплина сознательно ограничена рамками «социальной информатики» или профилирующих учебных предметов документально-комуникационного цикла.

Понятие «информатика» (от англ. Information science) трактуется часто значительно шире. Зарубежные ученые вкладывают в ее определение следующее содержание:

- «комплексная наука, которая изучает свойства и поведение информации, силы, управляющие потоком и использованием информации – как ручные, так и машинные с целью оптимального хранения, поиска и распространения…» и далее —
- «… способы представления информации как в естественных, так и в искусственных системах; изучение таких устройств и методов обработки информации, как вычислительные машины и системы программирования».

Термин «информатика» не имеет единого определения и на других языках мира. Так, в немецком языке – это наука о документации (dokumentations - wissenshaft); в английском – наука об информации (information science); в русском – информатика. Система терминов для обозначения дисциплины («информатика»), деятельности («научно-информационная деятельность»), объекта («научная информация») была введена в русском языке, как уже было отмечено выше, учеными ВИНИТИ в 1966 г.

Предметом изучения информатики являются:

- Структура научной информации
- Общие свойства научной информации
- Закономерности всех процессов научной коммуникации (формальных и неформальных)

Общим методом исследования в информатике, как и в других. науках, является диалектический метод, а для отдельных сторон научно-информационной деятельности применяются те же методы, как и в других отраслях науки. Информатика не имеет частных методов исследования, которые свойственны только ей одной.

Широкая область интересов информатики обусловило различные толкования ее объекта и предмета изучения. И поскольку к настоящему времени не сложилось единое

общепринятое определение информатики как науки, можно говорить о концепциях информатики как науки, которые развиты отечественными и зарубежными учеными.

3. Основные концепции информатики.

Отечественные трактовки информатики (Украины и СССР)

Рассмотрим основные отечественные концепции. (Подробная характеристика концепций дана в учебнике Хохловой Н.В., С.10 (табл. 1.1 Основные отечественные концепции информатики как науки)).

Итак, одна из концепций – Информатика, авторами которой были ученые МГИК (Московского государственного института культуры). В учебнике под ред. Тараканова К.В. дается следующее определение информатики – «информатика как наука изучает закономерности информационных процессов в социальных коммуникациях». Авторы учебника, коллектив сотрудников МГИК, определяют объект информатики следующим образом— информация как таковая, предмет – информационные процессы в искусственных системах, т.е. процессы сбора, хранения, обработки, поиска, выдачи и доведения до потребителя информации, пути организации и способы управления процессами, а также закономерности влияния процессов переработки на характер социальных коммуникаций. Главное внимание уделено закономерностям информационных процессов в искусственных информационных систем и три аспекта этих закономерностей: технология процесса, прогноз его развития и управление процессом. Соответственно авторами и выдвигается задача информатики – изучить закономерности сбора, хранения, обработки, поиска, выдачи и доведения до потребителя информации как таковой, пути организации и способы управления этими процессами, а также закономерности влияния процессов переработки на характер социальных коммуникаций.

Еще раньше – в 60-е годы была разработана авторитетная и признанная в стране концепция – Научная информатика. Основополагающие понятия этой науки (длительное время доминирующей в стране) были сформированы ведущими работниками ВИНИТИ Михайловым А.И., Черным А.И. и Гиляревским Р.С. на основе изучения зарубежного опыта, и изложены в монографиях, многочисленных публикациях и выступлениях. Суть этой концепции заключается в том, что информатика трактуется как научная дисциплина, изучающая структуру и общие свойства научной информации, а также закономерности всех процессов научной коммуникации. При этом подчеркивается, что «информатика – это пока лишь научная дисциплина, а не самостоятельная отрасль науки; информатика изучает структуру и общие свойства научной информации, а не любой информации и даже не семантической информации; информатика занимается изучением всех процессов

научной коммуникации, осуществляемых как по формальным каналам (т.е. через литературу), так и по неформальным каналам (личные кантакты между учеными и специалистами, переписка, обмен препринтами и т. д.); информатика относится к кругу общественных дисциплин, т.к. она занимается изучением явлений и закономерностей, свойственных лишь человеческому обществу» (3, с. 394-395)

Таким образом, объектом научной информатики является научная информация (логическая структура знания) и закономерности научно-информационной деятельности (ее теория, история, методика, организация), которая заключается в сборе, обработке, хранении, поиске и распространении научно-технической информации (НТИ).

Автоматизация процесса общественной коммуникации потребовала от представителей этой концепции уточнения предмета и объекта информатики. В частности, Горькова В.И. (ВИНИТИ) предложила считать информатику отраслью науки, изучающей процессы сбора, передачи, переработки, хранения, поиска, распространения и использования научной информации разрабатывающей методы и средства реализации этих процессов с помощью информационной техники в целях повышения эффективности системы общественной коммуникации.

Наиболее актуальные проблемы, изучением которой занимается научная информатика, следующие: семиотические[1] основы научной информации и научной коммуникации, моделирование информационных процессов и систем; свойства и закономерности документальных информационных потоков, автоматизация семантической (смысловой) обработки информации (автоматизация реферирования, индексирования, перевода); автоматизация информационного поиска на базе современной вычислительной техники.

Концепция – Социальная информатика развиваемая в ЛГИК (Ленинградский государственный институт культуры) с начала 70-х г.г. под руководством Соколова А.В. Согласно этой концепции, информатика понимается как общественная наука, изучающая общие структуру и свойства социальной информации и общие закономерности информационного обслуживания в обществе. Концепция была изложена в учебных пособиях, научных публикациях, и в настоящее время имеет свое продолжение.

Объект социальной информатики – социальная информация и общие закономерности социальных коммуникаций. Предмет – общие принципы, методы, закономерности информационного обслуживания, действительные во всех видах социальной коммуникации, как массовых, так и специальных, документальных и не документальных. В концепции подчеркивается, что формирование обобщающей теории не отрицает целесообразности существования «специальных» видов информатики

[1] Семиотика – наука, изучающая общие свойства знаков и знаковых систем.

(научной, экономической, патентной и др.), которые представляют собой частные прикладные теории. Объектом изучения «специальных» видов информатики являются различные типы социальных коммуникаций, а предметом – закономерности информационного обслуживания в рамках той или иной системы.

Еще одна концепция – Документалистика – предложенная Научным Советом по кибернетике при Президиуме АН СССР. Трактуется как одна из прикладных отраслей кибернетики, занимающаяся оптимизацией управления документальными системами всех типов – от избирательного искусства до делопроизводства. Объектом изучения является документальная коммуникация, предметом – закономерности функционирования документальных систем. При этом под документом понимается любой материальный носитель семантической информации, которая может быть выражена любой знаковой формой и зафиксирована любым образом. Т.о., документалистика изучает все взаимосвязи и явления, имеющие отношение к документам и документальным системам всех типов. Особое внимание уделялось механизации и автоматизации процессов накопления, поиска, распространения всех типов документов (с помощью перфокарт и микрофотокопирования). В настоящее время основные идеи этой концепции нашли отражение в учебных дисциплинах, связанных с документом (документоведение, документология, документные ресурсы и пр.)

Все выше перечисленные концепции объединяет то, что своей основной задачей авторы видят в исследовании путей оптимизации информационного обслуживания общества; сама информатика представлена общественной дисциплиной по изучению закономерностей социальной коммуникации в целом и отдельных ее видов (научной, документальной), а проблематика исследований информатики касается, прежде всего, самого феномена информации и информационных носителей, документальных информационных потоков информационных систем, проблем автоматизации информационного поиска и пр.

Другое направление концепций связано с представлением о том, что именно прогресс в вычислительной технике является определяющим для развития всей сферы научной и социальной информатики. Под информатикой стали понимать совокупность научных дисциплин и средств обработки информации с помощью вычислительной техники, а также сферу применения вычислительной техники в различных областях человеческой деятельности.

Так, накоплены точки зрения ученых (акад. Ершова А.П., Велихова Е.П., Михалевича В.С., Дородницина А.А.) на определение информатики как Прикладной (технической) информатики (например, как фундаментальная естественная наука, отрасль

народного хозяйства, инженерную дисциплину и пр.). В рамках этой концепции термин «информатика» понимается двояко:

- как наука, изучающая процессы передачи и обработки информации в системах технической коммуникации и
- как «сумма технологий», объединяющая технические средства в виде вычислительной техники и техники связи, программное обеспечение и соответствующие логические и математические методы.

В рамках прикладной информатики выделяют еще две разновидности: Безбумажная и «программистская» информатика. Концепция безбумажной информатики была развита сотрудниками института кибернетики им. В.М. Глушкова АН Украины, Глушков В.М., Михалевич В.С., Каныгин Ю.М. и др., предмет ее изучения – закономерности кибернетизации технических коммуникаций. «Программистская» концепция была развита школой акад. А.П. Ершова и представителями Отделения информатики, вычислительной техники и автоматизации АН СССР, Института проблем информатики АН СССР. Ее предмет — закономерности математического моделирования на ЭВМ процессов обработки и передачи информации. Объединяет эти две концепции один объект изучения — техническая коммуникация.

Анализируя существующие толкования, следует подчеркнуть, что все концепции и трактовки информатики как науки можно сгруппировать в два блока научных дисциплин:

1. Социальные (общественные) и
2. Технические.

Обобщив концепции первого блока, остановимся на таком определении ***информатики*** *(для гуманитарной сферы)* ***— это научная дисциплина, изучающая структуру и свойства информации, закономерности информационных процессов, методы и средства их реализации в системе коммуникаций (социальных и технических) с помощью информационной техники (которая включает технику компьютерную, телекоммуникационную, связи, копировально-множительную и пр.).***

Информатика содержит две части (или аспекта изучения): теоретическую и прикладную (практическую). В теоретическом разрезе информатика выявляет и формулирует внутренние устойчивые зависимости и закономерности научной информации, в прикладном аспекте – разрабатывает рекомендации по оптимизации информационного сервиса, использованию информационных ресурсов и информационных технологий в целом. В соответствии с этими аспектами в высших учебных заведениях, готовящих специалистов в области информатики, преподаются две

учебные дисциплины: Теоретическая информатика и Информационные технологии (прикладная информатика).

Следует отметить еще одну научную точку зрения на информатику, которая принадлежит известному философу А.Д. Урсулу. Он который предложил основными значениями термина “информатика” считать следующие:

- Совокупность средств автоматизированной техники и технологии;
- Особая инфраструктурная область народного хозяйства, включающая вся сферу автоматизированной обработки и технологического использования информации;
- Отрасль научного знания, изучающая процессы передачи информации и средства ее автоматизированной переработки;
- Теория научной информации и научно-информационной деятельности.

Зарубежные трактовки информатики и ее предмета.

За рубежом имеется много трактовок информатики, взаимоисключающих и дополняющих друг друга. В качестве примера можно привести наиболее глубокие и оригинальные трактовки, получившие признание в Японии, Германии, Франции, США.

В Японии информатикой называется одна из четырех областей использования ЭВМ (направлений компьютеризации) – компьютеризация социально-коммуникативных процессов (связей между людьми и коллективами).

Информатика как наука в Японии, Германии, Франции, США относится к так называемым «социально-информатическим учениям». Так ,Французская Академия наук определила, что информатика – это наука о рациональной обработке – при помощи автоматических устройств информации, рассматриваемой как носитель человеческих знаний и связей в технологических, экономических и социальных областях.

В Германии признание получили положения, согласно которым предметом информатики являются информационные процессы и информационные системы, обеспечивающие коммуникацию между людьми. В качестве критерия ограничения используется понятие «социальная информация». В предметную область информатики включают информационные системы, научные, технологические, плановые и административные, политические, вспомогательные (системы кредитной информации, информации биржи труда и др.), юридические, городские, справочные и др., т.е. специальные информационные системы.

В США прочные позиции занял подход к информатике как науке о движении, переработке и использовании семантической информации в социальных системах, причем

особый акцент делается на технологических аспектах обработки и применения знаний; а центральное место занимает изучение технологических процессов сбора, кодирования преобразования, хранения, поиска, распространения, использования информации. причем технологические процессы изучаются не автономно, а в рамках информационных систем (сетей) с учетом документальных потоков и информационных потребностей. Информатика ориентирована на оптимизацию социальных информационных систем, обеспечивающих передачу семантической информации. В качестве средств и методов оптимизации информационных систем выступают ЭВМ, кибернетические и математические подходы (технология машинной обработки информации, математическое моделирование взаимодействия человека и машины и т.д.).

В проблематику информационной науки включается изучение структуры и свойств информации – социальной, машинной и биологической (частично).

В качестве областей практического применения результатов информационной науки рассматриваются различные массовые и социальные коммуникационные системы общества, в том числе школы и библиотеки, наука и экономическое управление, торговля, делопроизводство и телевидение, а также сложные человеко-машинные кибернетические системы.

Заслуживающие внимания трактовки информатики за рубежом связывают эту науку, во-первых, с общественной практикой, во-вторых, с информацией в виде человеческих знаний, в-третьих, с технологическими (т.е. машинизированными) способами переработки и использования информации.

4. Взаимосвязи с другими науками.

Информатика как самостоятельная научная дисциплина сформировалась на стыке многих наук: социальных, естественных, технических, поэтому она широко использует их методы и является интердисциплинарной (или междисциплинарной) наукой.

Взаимосвязь информатики с науками технического цикла, такими как математика, кибернетика, вычислительная техника, теория информации обнаруживается по многим вопросам, связанным с техническими средствами реализации информационных систем. Например, различные формы и методы обработки информации; информационные операции; средства информации, ее закономерности и многие другие.

Проблемы семантической информации, естественного языка как средств распространения информации одновременно с информатикой исследуются такими науками, как семиотика, лингвистика и др.

Информатика использует методы и результаты исследования многих других дисциплин: системного анализа, исследования операций, теории вероятностей, теории информации, теории передачи данных, науковедения, документалистики и др.

Из теории информации, теории передачи данных, семиотики информатика заимствовала понятие количества и качества информации, истолкование ее семантики и прагматики; использование статистических и алгоритмических подходов к феномену информации; к процессам создания структур информационных потоков с целью использования их технологии, разработки методов прогнозирования и управления, которые имеют непосредственное отношение к предметной области информатики.

Из документалистики восприняты подходы к документу как носителю информации, его информационные оценки.

Тесные взаимосвязи отмечаются между информатикой и дисциплинами библиотечно-библиографического цикла. Они обусловлены взаимосвязями соответствующих видов деятельности: научно-информационной и библиотечно-библиографической, которые включаются в систему социальных или информационных коммуникаций. Эти виды деятельности часто пересекаются или дополняют друг друга. Информатика заимствовала из библиотечно-библиографической деятельности: ряд процессов аналитико-синтетической обработки информации, дополнив их «своими» информационными операциями и методами; закономерности и свойства научных публикаций; результаты теоретических исследований в области библиотечно-библиографических классификаций для создания информационно-поисковых языков; формы и методы информационного обслуживания и др.

В то же время, информатика, опираясь на свои базовые научные дисциплины — вычислительную технику, кибернетику и др., разрабатывает свои собственные методы, исследует закономерности информационных процессов, связанные с обработкой информации компьютерными средствами.

Тема 2: Информация как объект изучения информатики.

1. Определения информации
2. Виды информации
3. Структура информации
4. Свойства информации

Список литературы:
[1 —3]

1. Определения информации.

Информация как вещество и энергия является основой окружающего человека мира (универсума).

Любая система, организованная определенным образом, содержит информацию. Чем более сложной является организация системы, тем больше информации аккумулируется в этой системе.

Обычно рассматриваются физические модели вещества и энергии, человеком созданной и человеком переданной информации. Однако информация, как таковая, существовала задолго до появления человека. Например, по мнению ученых, информация, закодированная в ДНК, появилась около 1 миллиарда лет тому назад.

До нашего века не было необходимости в исследовании информации в качестве отдельной сущности. Первыми, кто увидел такую потребность, были специалисты в области радио и телефонии. Так, британский ученый Хартли, занимаясь исследованием эффективности передачи, определил трактовку информации как независимой, абстрактной величины, и в 1928 г. получил уравнение для ее измерения.

Другая точка зрения ученых по вопросу: кто первым начал исследовать информацию? гласит следующее. Первое объяснение этого понятия дали журналисты и филологи в 20—30-е гг. 20-го века, они трактовали информацию как новости, сообщения, своеобразный газетный жанр, с помощью которого человек получает различного рода сведения.

Термин «информация» произошел от латинского слова informatio, которое переводится как изложение, разъяснение. Многие научные дисциплины стали использовать этот термин, хотя каждая из них вкладывает в него свое содержание (массовая, техническая, экономическая, медицинская, статистическая и пр.).

Научная трактовка термина «информация» связана с возникновением двух дисциплин:

- математической (или статистической) теории информации;
- кибернетики.

В теории информации американского инженера и математика Клода Шеннона (1948 г.), первоначально разрабатывавшейся применительно к случайным процессам и явлениям, для которых характерна неопределенность исхода, под информацией понимались не любые сведения, а лишь те, которые снимают полностью или уменьшают существующую до их получения неопределенность. Кратко можно сформулировать, что по теории К. Шеннона, информация — это снятая неопределенность. Пояснением этого тезиса будет следующее. Неопределенность существует тогда, когда может произойти одно из нескольких событий, т.е. система может перейти в одно из нескольких состояний. При этом количество информации, получаемой в результате снятия неопределенности, вычисляется по формуле, разработанной и носящей имя К. Шеннона.

Согласно этой формуле, за единицу измерения информации принята величина, названная ***битом*** (сокращенно от английского языка: ***bi****nary digi****t).*** (1 байт = 8 битов).

Теория К. Шеннона дала возможность количественного определения информации в сообщении. Однако она полностью игнорировала содержание передаваемой информации, оставляла в стороне смысл сообщения.

Учеными было сделано несколько попыток найти меру содержательности информации для ее получателя. Так, известны определения информации, рассматривающие ее с позиций:

а) теории отражения и разнообразия; и

б) теории отражения и управления.

Кратко рассмотрим суть этих подходов к определению информации.

Согласно первому подходу, информация трактовалась как отраженное разнообразие или разнообразие в отражении. Между отражением и информацией существует единство, т.к. когда мы говорим о передаче информации, то непременно связываем ее с процессом отражения одного материального объекта другим. Окружающий нас мир многолик в своем развитии: различны составляющие его материальные объекты, элементы, связи между ними, свойства, протекающие процессы. Там, где есть разнообразие, возникает информация (информационные процессы). Таким образом, информация выступает свойством, стороной отражения, тесно связанной с неоднородностью материального мира. Она содержит в себе отраженное разнообразие и характерна для всех форм и видов движения материи, в т.ч. и неживой природы.

Второй подход относят к кибернетико-семантической концепции, которая связала воедино информацию и отражение с понятием управления. Основоположником этой концепции считается американский ученый Норберт Винер, которые в 1948 г. предложил «информационное видение» кибернетики как «науки об управлении и связи в живых

организмах, обществе и машинах». В последние десятилетия 20-го века наметилась идея синтеза знаний о связи и управлении с так называемой «информационной теорией управления», развиваемой научной школой советского ученого Б.Н. Петрова. Согласно кибернетической теории информации, информация — не просто результат отражения. Она является обозначением содержания, полученного из внешнего мира. Информацию составляет та часть знания, которая используется для ориентирования, активного действия, управления, т.е. в целях сохранения качественной специфики, совершенствования и развития системы. Иными словами, информация — это действующая, «работающая» часть отражения знания. Авторы этих концепций — А.Н. Колмогоров, В.И. Кремянский, И.И. Гришкин и др. советские ученые.

Однако эта концепция не бесспорна. Она отрицает, в частности, существование информации в неживой природе.

Приведены лишь некоторые научные точки зрения, не исключающие и других подходов к понятию информация.

Информация, которая передается в человеческом обществе (социуме), и активно участвует в формировании общественного сознания, называют *социальной информацией*. Она определяется как отраженное разнообразие или концептуально связанные сведения, данные, понятия, отраженные в нашем сознании и изменяющие наши представления о реальном мире. Согласно такому определению, любой вид информации, функционирующей в обществе, относится к социальной информации: обыденная, массовая, эстетическая, научная и пр.

Обратимся к трактовке *научной информации* как объекта информатики. Воспользуемся определением, данным российскими учеными А.И. Михайловым, А.И. Черным и Р.С. Гиляревским — «научная информация — это получаемая в процессе познания логическая информация, которая адекватно отражает явления и законы природы, общества и мышления, и используется в общественно-исторической практике».

Информация может рассматриваться и как «сведения, являющиеся объектом хранения, передачи и/или преобразования». Рассмотрим, какие различия и связи выделяют между такими, казалось бы однородными, понятиями, как *знания, данные (сведения)* и *научная информация*.

В *знаниях* научная информация представлена в наиболее обобщенном и систематизированном виде и выражается в системах понятий, в суждениях, умозаключениях и теориях. Следовательно, научные знания — это не вся научная информация, а лишь ее некоторая часть.

Данные (или сведения) [data] — это информация, получаемая в процессе чувственного познания и еще не подвергнутая переработке и обобщению абстрактно-логическим мышлением человека. Данные, получаемые в процессе чувственного познания, никак нельзя считать научной информацией. Они служат лишь сырьем для ее создания.

По мнению зарубежных ученых, «информация — это данные, собранные и систематизированные в пригодную для использования форму». Согласно схематической трактовке американского политолога Ф. Хартли, «информация, рождаясь из «сырых» данных, достигая «зрелости», переходит в знания».

2. Виды информации.

Разработано много подходов к классификации информации, в основу которых положены различные ее признаки и особенности. Обобщив их, выделим лишь основные классификационные ряды информации, которые упоминаются большинством ученых в области информатики.

Основные виды информации выделяются по таким признакам:

1) по сфере возникновения:

- элементарная — возникшая в неживой природе;
- биологическая — возникшая в мире животных и растений;
- социальная — возникшая в человеческом обществе (социуме).

2) по способу передачи и восприятия:

- визуальная;
- аудильная (звуковая или фонетическая);
- аудиовизуальная;
- тактильная (осязательная);
- вкусовая.

3) по общественному назначению:

- массовая (общественно-политическая, обыденная, эстетическая и пр.);
- специальная (научная, производственная, техническая, управленческая и пр.);
- личная.

Кроме этих основных признаков, выделяли и другие: по типу передаваемой информации (документная, фактографическая и др.); по форме представления (текстовая, графическая и др.); по способу распространения (опубликованная, неопубликованная); по степени аналитико-синтетической переработки (первичная, вторичная); по области получения и/или использования (экологическая, историческая, географическая и пр.). Эти признаки или основания деления выделялись по отношению к научной информации;

возможны и другие варианты классификационных рядов по отношению к другим типам и видам информации.

3. Структура информации.

Структуру информации начали рассматривать на примере научной информации, позднее — по отношению к социальной информации. Признано, что структура научной информации имеет ярко выраженный иерархический характер и в ней можно выделить два аспекта (или структуры): содержательный и формальный.

Содержательную структуру информации ученые отождествили с научным знанием. Элементами 1-го уровня являются эмпирические факты о предметах, процессах или событиях реальной жизни. Этому уровню соответствует информация о научных фактах. На 2-м уровне — познании — происходит осмысление эмпирических фактов, устанавливаются между ними связи, эмпирический факт превращается в научный. Этому уровню соответствует информация о научных теориях и гипотезах, объясняющая и объединяющая некоторую совокупность научных фактов и взаимосвязь между ними. 3-й уровень, высший — уровень абстрактно-логического мышления, в процессе которого вырабатываются концепции и законы, образующие основы данной науки или области знания. Информация этого уровня объединяет некоторую совокупность научных, гипотез, концепций, теорий, законов. Выделение этих уровней в содержательной структуре информации в достаточной мере условна и не имеет явного выражения. В информационных сообщениях могут содержаться элементы каждого или любого уровня.

Формальная структура научной информации также иерархична, как и содержательная. Низшие уровни этой иерархии являются общими для всей семантической информации, в которой выделяют отдельные звуки, буквы, слова, фразы, смысловые комплексы, произведения, К высшим уровням иерархии относят научные документы.

Формальная структура предполагает анализ знаковой формы, с помощью которой передается содержание информации. Наиболее часто такой знаковой формой является естественный язык. Как известно, в структуре естественного языка выделяют три уровня: фонетический, лексический и синтаксический. Есть еще один уровень — текстовой, на котором выделяют логические синтагмы и текстовые парадигмы.

3. Свойства информации.

Информация характеризуется тремя категориями свойств: атрибутивными, прагматическими и динамическими. *Атрибутивные свойства* — необходимые свойства, без которых информация не может существовать. *Прагматические свойства* — свойства, характеризующие степень полезности информации для практики. *Динамические свойства* — свойства, характеризующие изменения информации во времени.

Дадим их характеристику.

1). Атрибутивные свойства включают следующие:

а) языковая природа: содержание информации может быть изложено на разных языках, от этого ее смысл не должен изменяться; т.е. характерна относительная независимость информации от языка и ее носителя;

б) дискретность: содержащиеся в информации конкретные знания характеризуют отдельные фактические данные, закономерности и свойства, которые распространяются в виде отдельных сообщений, состоящих из фраз, параграфов, глав и других фрагментов, объединенных в статьи, журналы и другие документы;

в) непрерывность: новая информация, зафиксированная в отдельных сообщениях, сливается с уже накопленной ранее, способствуя поступательному развитию научной мысли.

2). Прагматические свойства включают следующие:

а) наличие смысла и новизны: информация носит понятийный характер, т.е. именно понятия составляют смысл слов естественного языка, с помощью которого осуществляется коммуникация между людьми. Именно в понятиях обобщаются наиболее существенные признаки предметов, процессов и явлений окружающего мира; они представляют содержание информации, выражаемое в определенной знаковой форме;

б) ценность (полезность): считается, что информация является ценной, если она способствует достижению стоящей перед человеком цели. Это свойство присуще всем видам информации;

в) кумулятивность: свойство информации — накапливаться. Иными словами, с течением времени количество информации растет, она накапливается, происходит ее систематизация, оценка и обобщение накопленных знаний.

3). Динамические свойства включают следующие:

а) свойство роста информации заключается в способности ее увеличиваться в количественном отношении в соответствии с определенными закономерностями;

б) свойство старения информации заключается в уменьшении ее ценности с течением времени. Но следует помнить, что старит информацию не время, а появление новой информации, которая отвергает полностью или частично имеющуюся информацию, уточняет ее, дополняет, дает новое сочетание сведений, приводящее к получению дополнительного єффекта.

в) свойство рассеяния: способность информации рассеиваться по различным источникам. Подробнее динамически свойства информации будут рассмотрены в следующей лекции, поскольку эти свойства совпадают с закономерностями, характеризующими развитие научных публикаций.

Тема 3: Источники информации: документальные и фактографические

1. Документ как источник информации
2. Фактографические источники информации
3. Закономерности документально-информационных потоков.

Список литературы:
[1 —3]
4. Боднарский Б.С. Сущность и значение документации // Сов. библиогр.— 1937.— №1.— С. 41-50. (ХГНБК №671275)

1. Документ как источник информации.

Слово «документ» произошло от латинского слова “document[um]”— свидетельство (от слова doceo—учить, извещать). Первоначально это слово обозначало письменное подтверждение правовых отношений и событий.

В известном всем толковом словаре В.И. Даля документ определяется как «всякая важная деловая бумага; также диплом, свидетельство». Такое толкование очень близко к оригиналу.

Документ изучался в рамках документационного направления библиографии. Идея и термин документация принадлежат известному бельгийскому библиографу Полю Отле (конец 19 века); ему же принадлежит инициатива создания Международного института документации в г. Брюссель. Здесь же образовалась мощная Ассоциация международных ассоциаций с документационными при них аппаратами, положившими начало созданию Музея международной документации. В Брюсселе же, начиная с 1910 г. появляется целый ряд разнообразных специальных документационных институтов; был даже организован бельгийский СОюз документации. Документационное развитие отмечалось в Швейцарии (Цюрих), Франции (Париж) и некоторых других странах Западной Европы.

В России одним из первых обратился к изучению сущности и значения документации известный ученый Б.С. Боднарский [4]. Он дал следующее определение документа: «документ есть все, что графическими знаками изображает какой-либо факт или идею» [4, С.45]. Он раскрыл также две особенности документа: 1) большое

разнообразие и 2) главную их массу составляют «мелкие документы». Его группировка (или классификация) документов включила:

1) Простые документы: а) тексты: печатные издания и рукописи; б) идеограммы (рисунки, фотографии); г) условные обозначения.

2) Составные документы: а) репертуары; б) досье (доклады, протоколы, отчеты); в) атласы; г) энциклопедии.

Как видим, толкование термина «документ» многозначно и объясняется это многоаспектностью самого документа. Каждая научная дисциплина выделяет свой аспект в определении термина «документ».

В информатике применяется информационный подход к документу, который заключается в рассмотрении его в качестве разновидности информационного сообщения. Благодаря этому вскрываются функциональные и структурные свойства, присущие документу, как и другим носителям информации, создается возможность для содержательного и формального анализа документов с использованием семиотики, лингвистики, теории информации.

Обратимся к классификации информационных сообщений А.В. Соколова, в которой получили развитие более ранние классификации, в т.ч. и группировка Б.С. Боднарского.

I. Документальные сообщения:
   1. Кодированные:
      а) читаемые:
         -опубликованные;
         -неопубликованные.
      б) идеографические *(использующие языковые знаки — условные обозначения: географические карты, ноты, чертежи и пр.);*
      в) аудиальные *(звукозаписи —речи, музыки);*
      г) машиночитаемые.
   2. Некодированные:
      а) иконические *(несущие неязыковые знаки: рисунки, фотографии, диапозитивы, кинофильмы);*
      б) «документы» трех измерений *(вещественные объекты, выполняющие функцию знаков — музейные экспонаты, образцы пород, архитектурные памятники и пр.);*
      в) аудильные (*звукозаписи, кроме записи речи и музыки).*

II. Недокументальные сообщения:
    1. Межличностное неформальное общение.
    2. Формализованное общение, в т.ч. лекция, спектакль, радио- или телепрограмма.

Итак, согласно такой классификации, документ трактуется достаточно широко.

В настоящее время словом документ обозначают объекты, содержащие информацию на любом материальном носителе (бумаге, магнитном диске и пр.) при помощи какой-либо знаковой системы, которая предназначена для передачи во времени и пространстве. В таком толковании документ — это не только письменное подтверждение, но и все публикации, и закрепленные на любых носителях сведения, используемые во всех областях деятельности человека.

Документальные источники информации подразделяются на первичные и вторичные. Первичные документы — это документы, содержащие исходную информацию. Вторичные документы — это документы, являющиеся результатом аналитико-синтетической переработки одного или нескольких первичных документов. Первичные документы и их совокупность включает опубликованные, неопубликованные и непубликуемые документы. Каналами распространения первичных документов служат различные виды изданий: периодические и непериодические; научные, учебные, массово-политические, научно-популярные, производственные и другие. Вторичные документы публикуются в информационных изданиях и также подразделяются на различные виды. По характеру включений информации и целевому назначению информационные издания подразделяют на библиографические, реферативные и обзорные.

Подробнее о классификациях и разновидностях документов можно узнать из учебных курсов по документоведению и библиографии.

2. Фактографические источники информации.

Фактографические источники информации содержат информацию в виде конкретных фактов, фактических событий или их совокупности, зафиксированных на каком-либо материальном носителе.

Между фактографическими и документальными источниками информации (или информационными сообщениями) имеется много общего. И те, и другие зафиксированы на материальном носителе, перемещаются во времени и пространстве, бывают простыми и сложными, повторяющимися и неповторяющимися. Между ними есть и различия, например, один документ может содержать один или несколько фактов.

Фактографическая информация — это информация о фактах (в прошлом, настоящем, будущем) или их совокупности. Фактографическая информация представляется в виде

фактографических описаний, которые представляют собой перечень наименований признаков описываемого объекта и значений этих признаков.

Типичными носителями фактографической информации являются справочники, словари, промышленные каталоги, прейскуранты, статистические таблицы и пр.

Фактографические информационные сообщения создаются или путем извлечения фактов из документов, в результате аналитико-синтетической переработки, или путем непосредственной регистрации сведений о фактах на материальном носителе.

Фактографические сообщения содержат:

- результаты непосредственных наблюдений;
- фактографические изображения различных параметров материального мира;
- хронологические характеристики различных явлений;
- технико-экономические показатели, характеризующие параметры машин, оборудования, предметов;
- чертежи, схемы, рисунки и пр.;
- имена собственные (названия стран, городов, животных и пр.);
- физические, механические и математические модели;
- химические и математические формулы;
- теоретические положения и прогнозы в науке, технике, производстве; и др.

Среди фактографических информационных сообщений выделяют такую их разновидность, как — фактологические сообщения, которые создаются путем последующей логической переработки фактографических сообщений, и содержат факты, отсутствующие в явном виде в исходном положении.

3. Закономерности документально-информационных потоков.

Документально-информационные потоки характеризуются закономерностями, которые были выявлены и исследуются в области науковедения и информатики.

Поток информации — это множество сообщений (документальных и недокументальных), целенаправленно передающихся по информационному каналу. *Канал* — это путь, по которому информационное сообщение движется от отправителя к потребителю информации.

Наиболее существенными закономерностями, характеризующими развитие научных публикаций, признаны следующие:

- закономерность роста;
- закономерность старения;
- закономерность рассеяния.

Закономерность роста понимается следующим образом: совокупность документов постоянно и закономерно увеличивается как следствие роста, повторяемости и многократности использования накопленной обществом информации. Динамика роста информации определяется экспонентой.

Экспоненциальный рост объема накапливаемых в обществе документов был подтвержден многочисленными экспериментальными данными. Хорошо известен график Д. Прайса, характеризующий экспоненциальный рост числа названий журналов.

Принято функциональную зависимость динамики роста документально-информационных потоков оценивать периодом удвоения количества публикаций, находящихся в потоке. Количество научных публикаций, попадающих в мировой поток, в среднем удваивалось за 10-15 лет (интенсивность прироста k=0,7), к 1985 — каждые 5 лет, к 1990 — каждые 2 года.

Экспоненциальный рост числа публикаций характерен не только для традиционных, давно сложившихся наук, но и для новых, междисциплинарных научных направлений.

Закономерность старения понимается следующим образом: совокупность документов даже одного тематического направления разнородна не только по физическому и информационному объему, но и по ценности и полезности для потребителя закрепленной на них информации. В силу неодинаковой ценности разных документов в потоке скорость и характер их продвижения к потребителю по информационным каналам различны. Быстрее, как правило, находят своего потребителя документы, содержащие новую, и потому более ценную информацию. Документы обладают еще и свойством старения, когда спрос на них со временем падает.

Для измерения скорости старения публикации ученые Р. Бартон и Р. Кеблер предложили в 1960 г. принять меру, названную ими «периодом полужизни» (half-life) публикаций (впервые этот термин был предложен Дж. Берналом в 1958 г.) по аналогии с мерой, используемой для оценки скорости распада радиоактивных веществ. *Период полужизни публикаций* — это время, в течение которого была опубликована половина всех используемых в настоящее время изданий по какой-либо отрасли или предмету.

Например, если период полужизни публикаций по физике был равен 4,6 года, то это означает, что 50% всех ныне используемых (цитируемых) публикаций по этой отрасли имеют «возраст» не более 4,6 года.

Процесс старения публикаций описывается отрицательной экспонентой, а для ее аналитического выражения Р. Бартон и Р. Кеблер предложили формулу.

Закономерность рассеяния получила название «закон рассеяния Брэдфорда» по имени ученого, открывшего этот закон в 1934 г. — С. Брэдфорда, американского химика и

библиографа. Закон был полностью сформулирован лишь в 1948 г. Его краткая формулировка: “Если научные журналы расположить в порядке убывания числа помещенных в них статей по какому-либо заданному предмету, то в полученном списке можно выделить ядро журналов, посвященных этому предмету и несколько групп или зон, каждая из которых содержит столько же статей, что и ядро. Тогда числа журналов в ядре и последующих зонах будут относиться как 1: n : n²”. Закон Брэдфорда можно выразить уравнением, диаграммой и графическим сопровождением.

Позднее эту закономерность дополняли и уточняли другие исследователи. Советский исследователь Л.С. Козачков назвал ее закономерностью концентрации и рассеяния, так как наряду с рассеянием основная часть публикаций по теме концентрируется в сравнительно небольшом числе профильных изданий (ядро).

Феномен рассеяния публикаций необходимо знать и учитывать как ученым и специалистам, так и персоналу, занятому их информационным обслуживанием для достижения полноты, оперативности и комфортности.

Тема 4: Информационные процессы

1. Понятие об информационных процессах.
2. Сбор информации.
3. Обработка информации.
4. Хранение информации.
5. Поиск информации.
6. Распространение информации.

Список литературы:
[1 —3]

1. Понятие об информационных процессах.

Информационным процессом называется взаимодействие между сообщением и отправителем и потребителем информации.

Иными словами, информационные процессы — это совокупность последовательных операций, действий и связей по обмену информацией, осуществляемых в системе коммуникаций.

В соответствии с каналами связи различают *информационные процессы: формальные и неформальные*. К *неформальным* относят процессы, которые выполняются непосредственно самими учеными или специалистами: диалог между ними, посещение научно-производственных подразделений и лабораторий; выставок; обмен письмами, публикациями. Для неформальных процессов характерно то, что в коммуникациях обязательное участие принимают сами ученые или специалисты; и информационные

процессы неотделимы от их профессиональной деятельности. *Формальные процессы* сформировались постепенно в ходе специализации, общественного разделения труда и получили свое организационное оформление, которое проявилось в таких сферах деятельности, как: редакционно-издательская, книготорговая, библиотечно-библиографическая, архивное дело и др.

Особое место принадлежит научно-информационной деятельности (НИД). В понятие НИД входят следующие взаимосвязанные и взаимообусловленные информационные процессы: сбор; аналитико-синтетическая переработка (преобразование); хранение; поиск; распространение.

2. Сбор информации.

Сбор информации — это процесс, с которого начинается вся информационная работа. Он заключается в получении информационными службами сообщений всех видов по различным каналам связи. Этот начальный процесс — важнейший для всех последующих информационных процессов, для информационной деятельности в целом.

Понятие "сбор" и "комплектование" иногда используют как синонимы. Однако сбор документов и информации понимается как составная часть более широкого понятия комплектования. Комплектование применяется в основном для обозначения комплектования фондов первоисточниками (или документами).Если же в системе информационного обслуживания происходит комплектование вторичными источниками и информацией ,то чаще всего используются термин "сбор" т.е., в информационно-библиотечной практике принято использовать эти термины для определения одного и того же процесса — библиотеки комплектуются первичными и вторичными документами, а службы информации собирают документы и информацию для формирования информационных массивов.

Информационные сообщения, зафиксированные в документах и на других носителях информации, как уже было сказано, собираются в фонде или массиве информации.

3. Обработка информации.

Следующий за процессом сбора информации — процесс обработки информации, который разделяется на обработку: 1) техническую и 2) научную. Техническая обработка заключается в учете и регистрации поступивших сообщений, проверке их на дублетность с имеющимися в фонде. Научная обработка заключается в информационном анализе и синтезе сообщений; и иначе называется аналитико-синтетической обработкой (или переработкой) информации.

Информационный анализ — это начальный этап преобразования документальной информации, состоящий в изучении документов и извлечении из них наиболее

существенных сведений. Этот процесс неотделим от синтеза, т. е. обобщения информации, полученной в результате информационного анализа документов, и подготовки результатов обобщения в той или иной форме.

Т. о. аналитико-синтетическая переработка — это совокупность процессов преобразования содержания документов с целью их анализа, извлечения необходимых сведений, а также оценки, сопоставления и обобщения. В процессе аналитико-синтетической переработки происходит свертывание информации, задачи которого заключаются в содержательной оценке социальной значимости документа (его научной, культурной ценности) и уменьшении физического объема при условии минимальной потери информативности.

К процессам аналитико-синтетической переработки относятся такие процессы:

- библиографическое описание;
- аннотирование;
- систематизация;
- предметизация —

— *эти операции выполняются обычно в библиотеках;*

- реферирование;
- индексирование;
- научный перевод;
- составление обзоров;
- извлечение из документов фактов —

— *эти операции выполняются как правило в службах информации.*

Дадим определения основным процессам аналитико-синтетической переработки документов:

- Составление библиографического описания – процесс стандартизованного представления основных данных о документе;
- Аннотирование – процесс составления кратких сведений, характеризующих документ со стороны его содержания, направленности, ценности, назначения, оформления и происхождения (аннотация отвечает на вопрос: *о чем?*);
- Реферирование – процесс составления реферата, представляющего собой краткое изложение содержания документа, методики исследования и его результатов, а также времени и места проведения исследований (реферат отвечает на вопрос*: что?*);
- Составление обзора – процесс свёртывания информации о множестве документов; иными словами можно сказать, что в обзорах обобщаются сведения, содержащиеся в различных документах. Различают три типа обзоров: библиографический; реферативный

(характеризуется большой глубиной обобщений); аналитический (результат анализа документальных источников и оценки; определённой интерпретации извлечённых из них фактов; их в свою очередь подразделяют на итоговые, прогностические, обзоры-обоснования);

- Индексирование – это процесс описания содержания и формы документа средствами искусственного информационно-поискового языка (т.е. выбор ключевых слов и перевод на ИПЯ (информационно-поисковый язык) или замена их соответствующими лексическими единицами (индексами));
- Научный перевод – это перевод текстов (научных, технических, экономических и др.) с одного языка на другой;
- Извлечение фактов из документов – это такие факты, как технические характеристики, свойства веществ и материалов, демографические характеристики и др.

4. Хранение информации.

Хранение информации — это процесс, связанный с обеспечением сохранности собранных и обработанных (в информационных службах) сообщений для передачи их в пространстве и времени.

Информационные сообщения, реализованные в определенной материальной форме, могут храниться в службах:

- документальной информации (книгохранилищах, депозитариях, библиотеках, архивах, музеях и т.п.);
- фактографической информации (редакция газет, телевидения, адресных и справочных бюро и т.п.);
- концептографической информации (службах патентной экспертизы, прогнозирования);
- комплексных информационных службах (службах и центрах информации).

5. Поиск информации.

Информационный поиск понимается как совокупность логических и технических операций, имеющих конечной целью нахождение документов, сведений о них, фактов, данных, релевантных запросу потребителя.

В зависимости от искомого объекта и цели различают два вида информационного поиска:

- документальный поиск, т. е. поиск сведений о документе (библиографическое описание, аннотация, реферат) или собственно документа (первоисточника или его копии);

- фактографический поиск, т. е. поиск данных, фактов, извлеченных из документов или функционирующих отдельно (характеристики приборов, свойств, материалов).

Важному информационному процессу — поиску — посвящена следующая 5 тема.

6. Распространение информации.

Распространение информации – это завершающий информационный процесс, суть которого заключается в выдаче ответа на запрос потребителя.

Различают два основных режима распространения информации (или информирования): справочный и текущий. Справочный режим предполагает доведение до потребителя ретроспективной информации, в ответ на разовый запрос. Текущее информирование заключается в предоставлении потребителям информации о новых поступлениях в систему и осуществляется массовыми, групповыми и индивидуальными методами, хорошо известными в практике информационного обслуживания.

Избирательное распространение информации (ИРИ) – одна из наиболее часто применяемых форм текущего информирования, позволяющая оперативно, систематически и дифференцированно удовлетворять информационные потребности специалистов в соответствии с их постоянными запросами. Абонементами системы ИРИ могут быть как индивидуальные, так и коллективные потребители. В отечественной информационной практике накоплены следующие разновидности системы ИРИ — системы дифференцированного обслуживания руководителей (ДОР); тематического обслуживания руководителей (ТОР); проблемно-ориентированного информирования руководителей. Эти системы отличались глубиной анализа предоставляемой потребителю информации и наличием обратной связи с потребителем информации.

Тема 5: Информационный поиск. Информационно-поисковые системы (ИПС) и их компоненты.

1. Информационный поиск: основные понятия и определения.
2. Виды информационного поиска.
3. Стратегия поиска и критерии выдачи информации.
4. Эффективность информационного поиска, основные показатели.
5. ИПС как средство реализации информационного поиска: общие понятия.
   5.1. Основные компоненты ИПС.
   5.2. Классификация ИПС.

Список литературы:

[1 —3]
4. Соколов А.В. Информационно-поисковые системы: Учеб. пособие для вузов. — М.: Радио и связь, 1081. — С.8—40; 66—70.

1. Информационный поиск: основные понятия и определения.

Термин «информационный поиск» впервые ввел в научный обиход ученый К. Муэрс. Он понимал его как процесс поиска и выдачи информации в соответствии с ее тематическим содержанием. Позднее ввели более широкое понимание информационного поиска.

Информационный поиск — это процесс нахождения в определенном упорядоченном множестве сообщений тех, которые соответствуют запросам потребителя или содержат необходимые потребителю факты, данные. Иными словами, информационный поиск понимается как совокупность логических и технических операций, имеющих конечной целью нахождение документов, сведений о них, фактов, данных, релевантных[2] запросу потребителя.

В понятийном аппарате информатики часто используют такие термины, как **«релевантность»** и **«пертинентность».**

Релевантной информацией называется такая информация, содержание которой соответствует полученному запросу.

Пертинентной информацией называется такая информация, содержание которой соответствует информационной потребности.

Следует пояснить в таком контексте суть таких понятий, как “информационная потребность“, “информационный интерес“, “информационный запрос“.

Информационная потребность — потребность отдельного человека, коллектива, общества в знаниях.

[2] Релевантность — это степень соответствия искомого документа запросу потребителя.

Информационный интерес — это осознанная информационная потребность. Объективно существующие информационные потребности отражаются в человеческом сознании и осознаются в виде информационных интересов.

Информационный запрос — это форма, в которой выражается информационный интерес. Информационный запрос, с которым потребитель обращается в информационную службу — это словесная формулировка информационной потребности. Все эти понятия активно используются при выполнении информационного поиска и особенно— при определении его эффективности.

Информационный поиск реализуется с помощью информационно-поисковых систем (ИПС).

Объектом информационного поиска могут быть документы (первичные и вторичные), фактографические описания в целом или в виде отдельных фрагментов. Другими словами, объектом информационного поиска может быть как материальный объект, так и описание этого объекта.

Целью информационного поиска может быть нахождение самого документа, данных о наличии или местонахождении документа либо разыскание элементов знания (смысловых элементов) с последующим их представлением пользователю. Конечная цель поиска бывает различной в зависимости от характера запроса.

2. Виды информационного поиска.

В зависимости от цели различают информационный поиск *адресный и семантический*.

Адресный поиск — это процесс разыскания информационных сообщений по чисто формальным признакам, указанным в запросе пользователя. Для осуществления адресного поиска необходимы следующие условия: наличие точного адреса и обеспечение строго определенного порядка расположения этого сообщения в ИПС. Адресами информационных сообщений могут выступать элементы библиографического описания документов, их авторские знаки, инвентарные номера и пр. Строго определенный порядок расположения информационных сообщений может быть обеспечен с помощью алфавита или путем возрастания последовательности нумерации.

Семантический поиск — это процесс разыскания информационных сообщений по их смыслу, содержанию. Одним из условий осуществления семантического поиска является перевод содержания информационных сообщений (и запросов) с естественного языка на информационно-поисковый язык (ИПЯ), формализованный. (Вопросы ИПЯ будут рассматриваться позже, в отдельной теме).

Принципиальная разница между адресным и семантическим поиском состоит в том, что при адресном поиске информационное сообщение рассматривается как объект, с точки зрения формы, а при семантическом поиске — как носитель знания, с точки зрения содержания. Цель адресного поиска — нахождение объекта или сведений о нем, семантического — нахождение информации по заданной теме, конкретному вопросу.

В зависимости от объекта различают информационный поиск: *документальный и фактографический.* Разновидности поиска по данному признаку представлены схематически на рис. 1.

Документальный поиск — понимается как процесс разыскания в ИПС первичных и вторичных источников информации, релевантных запросу потребителя. Различают два вида документального поиска: полнотекстовый (или библиотечный) и библиографический. Первый направлен на нахождение первичных документов (оригиналов или их копий). Второй — на нахождение вторичной информации, т.е. сведений о документах, представленных в виде библиографической записи.

Фактографический поиск — это процесс разыскания фактографической информации, релевантной информационному запросу. К фактографической информации относятся сведения, извлекаемые из документов или получаемые непосредственно от источников возникновения. Для обеспечения хранения и поиска эта информация фиксируется на специальных форматах в виде фактографических описаний — информационных сообщений, представляющих упорядоченную совокупность данных, признаков, характеристик, относящихся к некоторому предмету (процессу или явлению). Различают два вида фактографического поиска: документально-фактографический и фактологический. Первый заключается в отыскании в документах фрагментов текста, несущих фактографическую информацию или — заранее подготовленных фактографических описаний. Второй вид предполагает создание новых фактографических описаний в процессе поиска путем логической переработки имеющейся фактографической информации.

Различия между документальным и фактографическим поиском очевидны: в первом случае объектом поиска выступает документ, во втором — факт, отраженный либо в документе, либо в виде фактографического описания.

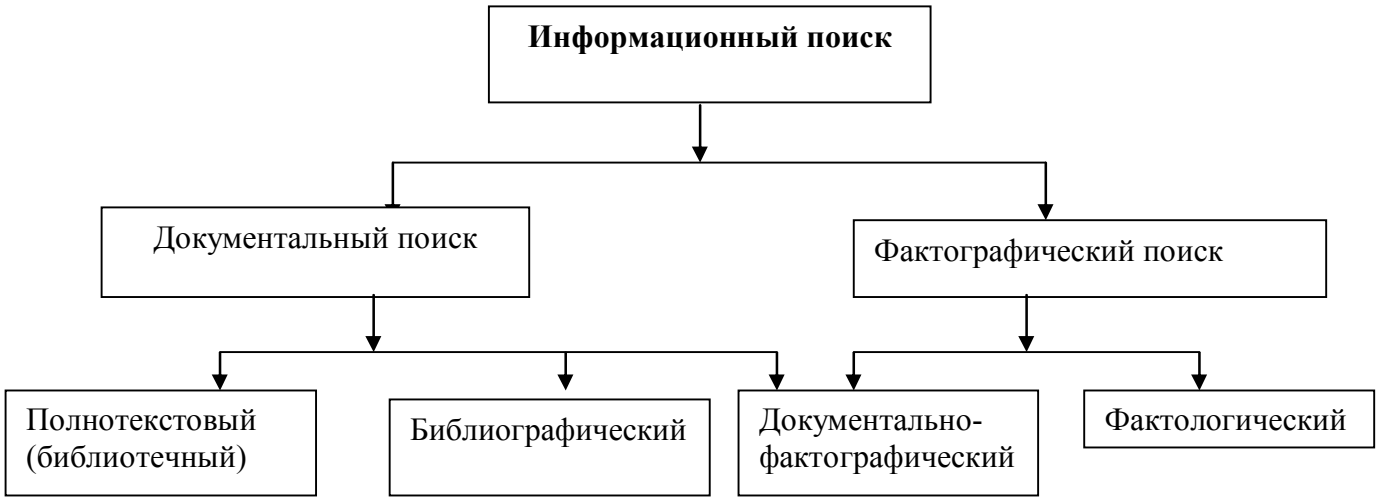

Рис. 1. Разновидности информационного поиска в зависимости от объекта

Цели и объекты информационного поиска взаимосвязаны, поэтому все виды поиска являются в основном пересекающимися. Например, полнотекстовый, библиографический, документально-фактографический поиск может быть как адресным, так и семантическим. Фактологический поиск — только семантическим.

Информационный поиск может осуществляться вручную или с помощью технических устройств, в зависимости от этого различают поиск ручной и автоматизированный.

3. Стратегия поиска и критерии выдачи информации.

Для осуществления всех видов информационного поиска необходимо иметь представление не только о цели и объекте поиска, но также и о правилах его проведения. Эти правила определяют *стратегию поиска*, т.е. способы оптимального достижения необходимых результатов — выдачу потребителю релевантной информации. Выработка стратегии поиска — это процесс творческий, который зависит от типа поисковой задачи, критериев выдачи, характера диалога между потребителем и ИПС. В информационной практике чаще всего встречаются такие типы поисковых задач:

- *адресный запрос* — о наличии в ИПС определенного информационного сообщения;
- *тематический запрос* — запрос на подбор определенного информационных сообщений или документов по определенной теме;
- *фактографический запрос* — требующий конкретного ответа, а не отсылки к документу;

- *запрос на уточнение данных* — требующий установления точного описания факта или документа.

Критерии выдачи — это формальные правила, по которым производится сопоставление поискового образа документа (ПОД) и поискового образа запроса (ПОЗ) и принимается решение о выдаче ответа на вопрос.

Поясним новые понятия: ПОД и ПОЗ.

ПОД (поисковый образ документа) — это набор индексов ИПЯ, соответствующих основным предметам и аспектам содержания документа, получаемый в результате индексирования документа.

ПОЗ (поисковый образ запроса) — некоторая совокупность понятий о главных темах запроса, выраженных на ИПЯ.

ПП (поисковое предписание), которое как правило сопровождает ПОЗ — это указания, необходимые для проведения информационного поиска.

Итак, согласно критериям выдачи определяется формальное соотношение между индексами ПОД и ПОЗ, независимо от их смыслового значения, т.е. сравнение смысла документа и запроса подменяется формально-логическими или вычислительными процедурами.

Различают критерии выдачи: теоретико-множественные и логические.

К теоретико-множественным критериям относятся следующие критерии:

- «на совпадение» — предполагает выдачу документов при полном совпадении индексов ПОД и ПОЗ;
- «на включение» — предполагает выдачу только тех документов, ПОД которых полностью включают индексы ПОЗ;
- «на пересечение» — предполагает выдачу тех документов, ПОД которых лишь частично совпадают с ПОЗ, т.е. сходны лишь некоторые индексы ПОД и ПОЗ,

Для повышения эффективности поиска задают его логику. С этой целью термины запроса связывают логическими связками или операторами. Различают следующие логические критерии выдачи:

1). Критерий *"логическая сумма"* — предполагает выдачу тех документов , в ПОД которых входит поисковый признак A, либо B, либо C и т.д. Запись поискового предписания в таком случае имеет следующий вид:

A *v* B *v* C *v* ....*v* Z, где *v* — знак логического сложения (оператор ИЛИ, OR, оператор дизъюнкции).

Логическое сложение поисковых признаков реализуется критерием «на полное совпадение».

2). Критерий *«логическое умножение»* — предполагает выдачу тех документов, в ПОД которых одновременно входят признаки A, B, C и т.д. Запись поискового предписания в данном случае имеет такой вид:

A ^ B ^ C ^….^ Z, где ^ — знак логического умножения (оператор И, AND, оператор конъюнкции).

Логическое умножение реализуется критериями «на включение» и «на пересечение».

3). Критерий «логическая разность» — предполагает выдачу тех документов, в ПОД которых входят признаки A, B, но не C. Символьно поисковое предписание записывается в таком виде:

A ^ B ┐ C, где ┐ — знак вычитания (оператор НЕ, NOT).

Возможен и сложный критерий, объединяющий два критерия, например, первый и второй или первый и третий.

Во всех случаях логические критерии определяют вид записи поискового предписания и задают формальную логику поиска сообщений в информационном массиве.

С позиции «общения» (диалога) потребителя с ИПС (локальной) поиск можно осуществлять тремя способами: по одному запросу в соответствии с индивидиальным ПП; параллельно по нескольким, предварительно накопленным запросам; с групповой обработкой нескольких предварительно накопленных и сгруппированных по критерию близости запросов. Схема поиска зависит от способа реализации ИПС и организации информационных массивов.

4. Эффективность информационного поиска, основные показатели.

Эффективность информационного поиска определяется рядом показателей, которые можно объединить в две группы: *семантические и технико-экономические*.

Семантические показатели включают показатели *полноты* и *точности* выдачи информации.

Полнота выдачи информации — это количественная характеристика информационного поиска, определяемая отношением между числом выданных на данный запрос релевантных документов и общим числом релевантных документов в информационном массиве.

Точность выдачи информации — это количественная характеристика информационного поиска, определяемая отношением между числом релевантных документов и общим количеством всех документов, выдаваемых в ответ на запрос.

Технико-экономические показатели — это *оперативность*, *трудоемкость*, *стоимость*.

Оперативность поиска — это среднее время выдачи ответа на запрос.

Трудоемкость поиска — затраты труда на поисковые операции при поиске одного документа.

Стоимость поиска — совокупность денежных и материальных затрат на информационный поиск одного документа/информационного сообщения.

Степень удовлетворения информационных потребностей характеризуется также прагматическими показателями, т.е. оценка полученных в результате информационного поиска сообщений осуществляется самими потребителями.

5. ИПС как средство реализации информационного поиска: общие понятия

Информационно-поисковые системы (ИПС) имеют различные определения, среди которых отметим основные, признаные и используемые в учебных посолбиях по информатике для гуманитариев. Прежде всего, специфика ИПС заключается в том, что такие информационные системы нацелены не реализацию двух основных процессов: сохранение информации и поиск информации. И в этом их основное отличие от других разновидностей информационных систем, например, интегральных информационных систем, которые объединяют в себе многие процессы.

5.1. Основные компоненты ИПС

Как средство реализации информационного поиска ИПС представляет собой совокупность методов и средств, предназначенных для хранения и поиска информации, документов, сведений о них, отдельных фактов, данных в соответствии с запросами потребителей информации.

Для выполнения любого вида информационного поиска в состав ИПС должны входить следующие компоненты:

1. информационный массив объектов — совокупность документов, фактов или их описаний;
2. логико-семантический аппарат, состоящий из ИПЯ, методов индексирования и поиска информации;
3. средства реализации, т.е. совокупность технических устройств и информационной техники, с помощью которых осуществляется хранение и поиск информации;
4. люди, взаимодействующие с системой: пользователи и обслуживающий персонал: инженеры, операторы и др. технический персонал.

Такое понимание ИПС дается в учебных пособиях Н.В. Хохловой и К.В. Тараканова. Несколько расширенный вариант составляющих ИПС дается в трактовке А.В. Соколова; он рассматривает аналогичным образом п.1, 3, 4, однако п.2 разбивает на три подпункта:

ИПЯ, правила индексирования документов и запросов, правила поиска документов с использованием критериев выдачи. Такое понимание ИПС совпадает в общих чертах с первой русскоязычной трактовкой ИПС, которая была дана известными российскими учеными в области информатики А.И. Михайловым, А.И. Черным, Р.С. Гиляревским — «ИПС в ее абстрактном виде понимается как совокупность ИПЯ (с правилами перевода с естественного языка на этот язык и наоборот) и критерия смыслового соответствия между поисковыми образами документов и поисковыми предписаниями».

ИПС — это некоторая совокупность или комплекс связанных друг с другом отдельных частей, предназначенных для выявления в каком-либо множестве элементов информации (документов, сведений и т.п.), которые отвечают на информационный запрос, предъявленный системе.

Общие понятия и определения ИПС сформировались в значительной степени на основе зарубежных трактовок, а также на основе многолетнего практического опыта библиотечно-поисковой работы, которая выполнялась на базе ручных аналогов ИПС — картотек и каталогов.

5.2. Классификация ИПС

При подходе к классификации ИПС исходят из требований пользователей к качеству функционирования системы, т.е. к ее способности выбирать из информационно-поискового массива требуемую информацию с достаточной полнотой, точностью и оперативностью. Такие требования учитываются как при разработке, так и эксплуатации ИПС.

Основными признаками, по которым классифицируются ИПС, с точки зрения удовлетворения требований пользователей, являются:

- тематика формирования информационно-поискового массива;
- вид и объект информационного поиска;
- режим функционирования;
- средства выполнения информационного поиска и другие.

ИПС можно классифицировать также по типу используемого ИПЯ, критериям выдачи ответа на запрос и др. признакам, хотя они считаются второстепенными с точки зрения интересов пользователей.

По тематике (или профилю) формирования информационно-поискового массива — ИПС подразделяют на универсальные, отраслевые, многоотраслевые (политематические) и узкотематические. Отраслевые ИПС создаются, как правило, центральными отраслевыми службами НТИ в соответствии с закрепленной за ними тематикой. Многоотраслевые ИПС создаются в региональных службах НТИ— межотраслевых

территориальных центрах НТИ, которых в настоящее время в Украине насчитывается 25. Узкотематические ИПС организуют в местных службах НТИ: на предприятиях, в организациях, в НИИ, в вузах и т.д.

Возможна также классификация ИПС по типам и видам документов, вводимых в информационный массив: патенты, промышленные каталоги, неопубликованные издания и пр., без учета их тематики. Так, по типам документов различают ИПС, хранящие документы: текстовые (фонды библиотек и других организаций); иконические (фототеки и пр.); идеографичесике (картотеки нот. технических решений); аудильные (фонотеки); машиночитаемые (базы и банки данных)

По виду и объекту информационного поиска — ИПС подразделяются на такие виды: документальные, фактографические и документально-фактографические.

Документальные ИПС предназначены для поиска документов в ответ на полученный запрос. В документальном поиске различают два вида: полнотекстовый (или библиотечный) и библиографический. Документальные библиографические ИПС обеспечивают хранение вторичных документов, а библиографический поиск осуществляется с целью нахождения данных о первичных документах и их адресов. Документальные ИПС (библиотечные или полнотекстовые) соответственно нацелены на поиск полных текстов документов или их копий. Следует отметить, что в реальных условиях документальный поиск чаще всего осуществляется в два этапа (или по двум контурам): поиск сведений **о документе (вторичных документов) и поиск первичных документов.**

Фактографические ИПС обеспечивают хранение и поиск фактографической информации в ответ на запрос аналогичного типа. Напомним, к фактографической информации принято относить как сведения, извлекаемые из документов, так и собственно описания фактов (или фактографические описания), т.е. сведения, получаемые непосредственно от источников их возникновения (специалистов, систем и пр.).

Основное отличие документальной ИПС от фактографической заключается в том, что в документальных ИПС единицами информации являются элементы библиографического описания, а также ключевые слова, фразы, термины, дескрипторы; а в фактографических ИПС единицы информации — это признаки и их значения (реквизиты). Это обязательные данные, котолрые устанавливаются нормативно-технической документацией: стандартами, техническими условиями. Реквизиты отражают определенные научные, технические, экономические свойства объектов, процессов, явлений и представляют собой логически неделимые элементы любой сложности;

описывают их количественные и качественные свойства. Совокупность признаков называют собщением об объекте. Каждое сообщение имеет свою определенную форму.

Документально-фактографические ИПС занимают промежуточное положение между двумя выше указанными видами ИПС. Результатом поиска в них являются запрашиваемые факты, сведения, данные со ссылкой на документ, в котором они зафиксированы. Различие между документальной и фактографической ИПС заключается лишь в объекте информационного поиска.

По режиму функционирования (или с точки зрения режима распространения информации) — различают три разновидности ИПС:

1. системы избирательного распространения информации (сокращенно ИРИ), которые обеспечивают периодический поиск информации (как правило 1 раз в две недели или в месяц) в информационном массиве новых поступлений в соответствии с постоянно действующими запросами и выдачу пользователям сведений о найденых документах или фактах. Для оптимизации работы этой ИПС между пользователем и системой устанавливается и поддерживается обратная связь. Такие системы, в недалеком прошлом (в середине 20-го века советского периода), будучи в ручном режиме, успешно реализовали свои возможности и не только как системы ИРИ, но и как другие системы текущего оповещения информацией, в зависимости от категорий пользователей, как системы: коллективного или массового текущего информирования.

2. системы ретроспективного поиска, которые осуществляют справочное обслуживание по разовым запросам в массиве информации долговременного пользования. Такой поиск иначе называют еще режимом «запрос—ответ».

3. «интегральные» системы, получившие такое условное название, нацелены на работу как в режиме текущего оповещения, так и в режиме справочного обслуживания.

Надо отметить, что и документальные, и фактографические ИПС могут работать в любых режимах распространения информации.

По средствам выполнения информационного поиска— ИПС подразделяются на ручные, механические и автоматизированные. Первые и вторые разновидности — это уже из истории развития ИПС, т.к. в настоящее время распостранены последние — автоматизированные.

К ИПС с ручным поиском относят каталоги, картотеки (в библиотеках, архивах и пр.), справочники, кроме того использовались карты с краевой перфорацией, унитерм-карты. В механизированных ИПС для поиска информации использовались счетно-перфорационные машины; для записи и хранения информации применялись перфокарты машинной сортировки.

В автоматизированных ИПС поиск информации реализуется на ЭВМ (или компьютерах): запись и хранение информации осуществляется на машиночитаемых носителях (магнитных и др. дисках); автоматизируется не только процесс поиска, но и процессы ввода, редактирования и актуализации, хранения, переработки и выдачи информации.

ИПС могут классифицироваться по многим другим признакам, которые не взаимоисключают друг друга.

Тема 6. Языки представления информации. Информационно-поисковые языки

1. Общие понятия естественных и искусственных языков.
2. Информационно-поисковые языки (ИПЯ).
   - Структура ИПЯ и требования к ним
   - Типы и виды ИПЯ
3. ИПЯ дескрипторного типа
   - Определения. Методика построения
   - Грамматика дескрипторных ИПЯ
4. Лингвистическое обеспечение информационных систем

Список литературы:

1. Хохлова и др. Информатика: Учеб. пособ. – М., 1990. – С.32-55
2. Соколов А.В. Информационно-поисковые системы: Учеб. пособ. – М., 1981. – С.41-86.
3. Информатика: Учеб. пособ./ Под ред. К.В. Тараканова. – М., 1986. – С.74-94.
4. Михайлов А.И., Черный А.И., Гиляревский Р.С. Основы информатики. – 2-е перераб. и доп. изд. – М., 1968. – С.316-515.
5. Гиляревский Р.С. Основы информатики: Курс лекций / Р.С. Гиляревский.— М.: Изд-во «Экзамен», 2003. — С.143-164.

1. Общие понятия естественных и искусственных языков.

Язык возникает в процессе развития общественного производства материальных благ. Мыслительная деятельность человека осуществляется в форме понятий и представлений, которые неразрывно связаны друг с другом и являются формой отражения реальной действительности в мышлении. И представления, и понятия, формируясь в сфере мышления, имеют внеязыковую природу, однако мышление с самого начала своего возникновения связано с языком. Процесс мышления совершается с помощью слов и предложений, являющихся единицами языка, которые представляют чувственную, материальную оболочку наших мыслей. Ни одна мысль не может возникнуть вне слова, именно слово сделало возможным переход от чувственных образов к суждению и понятию о вещах. От единичных предметов до самых абстрактных философских

категорий – все обозначается словом. Однако, как утверждают ученые, нельзя ставить знак тождества между языком и мыслью, которые хотя и неразрывно связанны друг с другом, являются различными общественными явлениями. Язык – это только орудие реализации мысли, обмена мыслями, возникающее в процессе социальных коммуникации.

Язык понимается как знаковая система любой физической природы, выполняющая познавательную и коммуникативную функцию в процессе человеческой деятельности.

Язык может быть **естественным и искусственным**.

Естественный язык понимается как «язык в собственном смысле, человеческий язык как орудие мысли и средство общения в отличие от его искусственных субститутов».[3]

В ходе многовековой общественной практики народы создают звуковые естественные языки, представляющие совокупность средств выражения содержания человеческого сознания (абстрактного мышления, чувственных образов, эмоций, воли и т.д.), способных передать в процессе коммуникации любую из его сторон.

Естественные языки как социальное явление имеют ряд функций:

- гносеологическую (познавательную) – с помощью языка обеспечивается сохранение опыта и знаний человека;
- номинативную – с помощью слов язык именует, называет вещи, явления, понятия;
- регулятивную – выделяется на основе коммуникативной и устанавливает отношения между участниками общения;
- эмотивную – состоит в выражении эмоций, чувств, переживаний, настроений;
- побудительную – связана с выражением требования, желания, просьбы, направленных на другого человека с целью побудить его к выполнению каких-либо действий;
- эстетическую – основана на способности слова действовать не только значением, но и оформленностью.

Эти и ряд других функций естественного языка очень связаны друг с другом.

Наряду с естественными языками в человеческом обществе большое распространение получили различные искусственные языки, к которым относится любой вспомогательный язык, созданный людьми для каких-либо узких целей. Для решения каких-либо задач в области науки и техники (машинные языки); для общения между людьми, говорящими на различных естественных языках (эсперанто и др.); между членами какой-либо ограниченной группы лиц (профессиональные диалекты). Искусственные языки создаются и существуют на базе естественных языков.

---

[3] Философский словарь. М., 1963. С.535.

Среди искусственных языков особое значение занимают *информационные языки*, создаваемые для лучшей реализации коммуникативной функции языка, т.е. обмена информацией. Необходимость создания и использования информационных языков для обработки информации возникла и продолжает углубляться в связи с совершенствованием информационных технологий в обществе. Широкое внедрение вычислительной (или компьютерной) техники в научно-информационной деятельности поставило вопрос об использовании *машинных языков (или языков программирования*). Это искусственные формальные языки, предназначенные для записи информации, хранящейся в запоминающем устройстве (ЗУ) вычислительных машин (компьютеров), для описания программ (алгоритмов), указывающих очередность арифметических и логических операций при решении той или иной задачи и последовательность выполнения команд по вводу данных из ЗУ, переработке и преобразованию поступающей в компьютер информации.

2. Информационно-поисковые языки (ИПЯ)

Для поиска информации разрабатываются и широко применяются *информационно-поисковые языки (ИПЯ)*. **ИПЯ** трактуется как искусственный язык, представляющий совокупность средств для описания формальной и содержательной структуры информации и ее поиска по запросу потребителей. Процесс представления информации на ИПЯ, в результате которого создаются поисковые образы (или признаки) документов (ПОД) и поисковые образы запросов (ПОЗ), называется индексированием.

Структура ИПЯ и требования к ним

Структура ИПЯ однотипна со структурой информации и предполагает выделение фонетического, лексического, синтаксического и текстового уровней языка, элементы которого объединяются в синтагмы и парадигмы. Внутреннюю структуру языка образует целостная, организованная, единая совокупность составляющих его сторон – фонетики, лексики, грамматики – и словообразования, частей этих сторон и отдельных лексических единиц (фонем, морфем, слов, словосочетаний, предложений и т.п.).

Поясним это подробнее.

В настоящее время существует большое количество различных ИПЯ, их комбинаций и модификаций. Особенно разнообразны лингвистические средства автоматизированных ИПС. Однако изучение специалистами различных ИПЯ способом сравнения показало возможность их единообразного описания и анализа. Для этого, по мнению А.В. Соколова, следует выделить основные структурные составляющие плана содержания ИПЯ, которые еще называются логико-лингвистическими универсалиями

информационных языков. Иными словами, можно выделить следующие универсальные структурные составляющие ИПЯ любого типа:

- лексические единицы;
- парадигматические отношения;
- синтагматические отношения.

Лексическая единица (индекс, слово на ИПЯ) – наименьшая осмысленная последовательность знаков, задаваемая при конструировании ИПЯ. Лексические единицы – это единицы смысла в ИПЯ. Они соответствуют отдельному слову или словосочетанию естественного языка или научному понятию. Совокупность лексических един7иц образует лексику ИПЯ.

Парадигматические отношения представляют собой внетекстовые смысловые отношения между лексическими отношениями ИПЯ, устанавливаемые на основании потребностей информационного поиска. Эти отношения учитывают сходство и различие в содержании лексических единиц, например, отношения: род-вид, целое-часть, предмет-свойство и др. На основе этих отношений лексические единицы группируются в парадигмы.

Синтагматические отношения представляют собой семантические отношения между лексическими единицами, входящими в один поисковый образ (текст на ИПЯ). Группа лексических единиц, связанных синтагматическими отношениями, образуют синтагму (фразу, предложение на ИПЯ).

Принципиальное различие между парадигматическими и синтагматическими отношениями заключается в том, что первые учитывают семантические отношения внетекстовой природы, т.е. те, которые не зависят от каких-либо текстов, а вторые выражают семантику контекста и зависят от нее. Одни и те же лексические единицы образуют различные по смыслу синтагмы.

Специалисты отмечают, что развитость структуры ИПЯ определяет его семантическую силу, т.е. имеющиеся в языке возможности полного и точного выражения результатов мышления (понятий, высказываний, умозаключений).

Рассмотрим детальнее парадигматические и синтагматические отношения в рамках анализа структуры ИПЯ.

Парадигматические отношения были определены выше как внетекстовые, объективно существующие смысловые отношения между лексическими единицами, которые устанавливаются и фиксируются в словаре языка исходя из потребностей информационного поиска. Эти отношения учитывают сходство или различия в объеме и

содержании лексических единиц и бывают ***сильными (логическими)*** или ***слабыми (ассоциативными).***

Поясним новые понятия.

Объем понятия - это множество предметов, которые охватывают данное конкретное понятие. Например, объем понятия «периодические издания» включает такие предметы, как газета и журнал. Количество предметов, входящих в объем понятия, может быть конечным (даже одно) и бесконечным (например, числа).

Содержание понятия - это отраженная в сознании человека совокупность свойств, признаков, присущих каждому предмету, входящему в объем понятия. В процессе познания содержание понятия может изменяться, т.е могут появляться новые признаки.

Сильные парадигматические отношения выявляются в результате сопоставления объема понятий. К ним относятся такие виды отношений:

1. равнозначности (эквивалентности)- между понятиями, объемы которых совпадают, но в содержании имеются различия (например, документ печатный – документ опубликованный);

2. подчинения (родовидовое) – такое отношение между понятиями, когда объем одного или нескольких понятий входит в объем другого (например, документы вторичные – родовое понятие, указатели, списки, рефераты – видовое понтяие);

3. соподчинения – между видовыми понятиями, в равной степени подчиненными одному родовому. Объемы их не совпадают, в содержании имеются как общие, так и отличительные признаки (например, книга, брошюра, листовка – виды непериодических изданий);

4. перекрещивания –между понятиями, содержание которых различно, но объемы частично совпадают (например, писатели и ученые);

5. противоположности – между соподчиненными понятиями, которые в своем содержании имеют несовместимые признаки, обуславливающие несовпадение объемов этих понятий;

6. противоречия – между двумя соподчиненными понятиями, видовые признаки которых несовместимы; это понятия, исключающие друг друга.

Слабые (ассоциативные) парадигматические отношения – выражают связи не между понятиями, а между самими предметами. При создании ИПЯ различных типов фиксируют в явном виде следующие *виды ассоциативных отношений:* целое - часть; система – элемент; отношение детерминации : причина – следствие; процесс – оборудование; процесс – материал; предмет – назначение; предмет – свойство; наука и объекты ее изучения; наука и ее представители и т.п.

Парадигматические отношения позволяют объединять лексические единицы ИПЯ в семантические группы – парадигмы, элементы которых обладают свойством взаимозаменяемости.

Кроме парадигматических в ИПЯ существуют ***синтагматические отношения*** *(синтаксические, грамматические, текстуальные)*, служащие для установления семантических связей между лексическими единицами. Средства выражения синтагматических отношений называют грамматикой ИПЯ. В языкознании грамматикой естественного языка называют объективно существующую в языке систему способов и средств построения и изменения слов и построения предложений. Составными частями грамматики естественного языка являются морфология и синтаксис. В информатике особое значение придается синтаксису ИПЯ, который и сводится к понятию грамматики.

Сравнение естественных языков и ИПЯ показывает, что между ними имеются существенные различия:

1. На *уровне лексики* – для включения в ИПЯ из всех частей речи допускаются только существительные. Другие части речи такие, как прилагательные, причастия, наречия, предлоги и союзы, используются только в составе словосочетаний. Местоимения, глаголы, деепричастия в состав лексики ИПЯ не включаются. В некоторых ИПЯ допускаются не только словосочетания – термины, но и словосочетания – предложения простейших типов. Таким образом, с лексическими единицами многих типов ИПЯ связаны грамматика, а иногда и фразеология естественного языка.

2. На уровне семантики лексических единиц – в естественных языках широко распространены неоднозначность т многозначность слов. Неоднозначность – это возможность выразить одну и ту же мысль разными словами – проявляется в наличии в языке синонимов (слов, отличающихся по звуковой форме, но совпадающие по значению); антонимов (слов, имеющих противоположное значение). Многозначность – свойство одинаково звучащих слов обозначать различное смысловое содержание – проявляется в полисемии (наличии у одного слова нескольких лексических значений: классификация (предмет), классификация (процесс)); омонимии (явлении, когда слова, имеющие одинаковую звуковую форму, различны по значении и происхождению: класс– общественная группа; класс-помещение, класс-разряд, Класс-коллектив и т.п.

При построении ИПЯ устраняют и синонимию, и полисемию, и омонимию.

Таким образом, главная отличительная особенность ИПЯ – простота лексики и грамматики по сравнению с теми же элементами естественного языка. Парадигматику любого языка можно отождествить с лексикой этого языка**,** а синтагматику – с грамматикой.

Требования к ИПЯ. Вследствие избыточности и недостаточности естественного языка он не может использоваться в качестве ИПЯ, т.к. это привело бы при поиске к большим потерям информации и к информационному шуму. Иными словами, значение слов естественного (повседневного) языка достаточно сложно, оно зависит не только от внешней формы слова, но и от обстоятельств, при котором оно высказано, иногда от субъективно-психологических факторов. Искусственный язык вводится для того, чтобы увеличить полноту и точность выдачи информации при поиске. Язык, который можно было бы применить в ИПС, должен быть формализован, т.е. в нем должна быть устранена многозначность слов естественного языка и все то, что характеризует отношение человека к разным предметам, эмоции, волевые побуждения. В таком языке должны быть выражены лишь объективные характеристики предметов и их соотношений.

Таким образом, можно выделить основные требования, которые предъявляются к ИПЯ.

1. *Однозначность* – каждая запись на ИПЯ должна иметь только один смысл и, наоборот, любой смысл должен получать единообразное представление на ИПЯ (отсутствие антонимов, синонимов и полисемичных слов).
2. *Эксплицитное (явное)* выражение полезных для поиска логических отношений и ассоциаций между словами ИПЯ.
3. *Открытость* ИПЯ, возможность его корректировки и дополнения.
4. *Удобство* пользования.

Однако эти требования невозможно выполнить полностью при разработке реальных ИПЯ, а построение идеальных ИПЯ практически невозможно. Поэтому реальным ИПЯ все же присущи и потери информации, и информационный шум при поиске информации.

Типы и виды ИПЯ (классификация ИПЯ)

К настоящему времени сформировалось большое количество типов и видов ИПЯ, и соответственно их классификаций. Наибольшое распространение получили следующие классификации: ВИНИТИ, Ф. Ланкастера и А.В. Соколова, которые различаются разными признаками. Классификация ИПЯ А.В. Соколова (ЛГИК) признана наиболее логичной и обобщающей предыдущие, в ее основу положены три видообразующих признака, учитывающих основные структурные элементы естественного языка: лексику, парадигматику и синтагматику. К этим признакам, другими словами, относятся способ задания лексических единиц (ЛЕ), способ координации (сочениания) ЛЕ и способ учета парадигматических отношений между ними. Учитывается и тип информационного поиска (ручной или автоматизированный), на который ориентирован тот или иной ИПЯ.

1. По признаку лексики (или способу задания лексических единиц) выделяют ***ИПЯ***: ***контролируемые и неконтролируемые.***

1.1. Контролируемые ИПЯ — это языки, лексика которых задается заранее с помощью словарей и таблиц. К ним относятся традиционные библиотечные классификации: УДК, ББК, фасетные классификации, язык предметных рубрик, ориентированные на ручной поиск, а также дескрипторный ИПЯ (словарь которого состоит из отдельных слов и словосочетаний, расположенных в алфавитном порядке), ориентированный на автоматизированный поиск.

1.2. Неконтролируемые ИПЯ — это языки, лексика которых не задается словарем, а строится на основе выбора неограниченного множества терминов естественного языка из индексируемых сообщений. Например, язык вспомогательных указателей, язык библиографического описания, предназначенные для ручного поиска; дескрипторные ИПЯ без контроля ПОД — для автоматизированного поиска.

2. По признаку синтагматических отношений (или способу координации (сочетания) лексических единиц) различают ***ИПЯ: некоординируемые и координируемые.***

2.1. Некоординируемые ИПЯ — языки, не допускающие координации своих ЛЕ ни в процессе индексирования (ПОД), ни в процессе поиска (ПОЗ). ИПЯ такого типа представляют собой жесткую классификационную схему, предусматривающую для каждого классифицируемого объекта одно и только одно определенное место, К ним относят, например, системы библиотечной расстановки книг, рубрикаторы информационных изданий, языки вспомогательных указателей.

2.2. Координируемые ИПЯ — это языки, в которых лексические единицы координируются между собой или в процессе индексирования, или в процессе поиска.

3. По признаку парадигматических отношений (или способу их учета) выделяют ***ИПЯ: иерархические, неиерархические и фасетные.***

3.1. Иерархические ИПЯ *(согласно А.В. Соколову)* — это языки, в которых все ЛЕ связаны сильными парадигматическими отношениями (подчинения и соподчинения) и образуют в совокупности иерархическую классификацию (т.е. «от общего к частному»).
Поясним основные понятия. Иерархическая классификация — это система классов, по которым распределяются понятия на основании наиболее существенных признаков, присущих этим понятиям и отличающих их друг от друга. Классификация — это процесс распределения понятий на взаимоисключающие классы.

В основе построения иерархических классификаций лежит деление понятий. Признак, по которому производится деление, называется основанием деления и является

переменным. Классификация строится таким образом, чтобы в получившейся системе каждый класс занимал относительно других классов определенное, точно зафиксированное место. Основной принцип деления понятий в иерархических классификациях — от общего к частному — основан на учете сильных парадигматических отношений, т.е. отношений подчинения и соподчинения.

Распространенная форма изображения иерархических классификаций — граф типа дерева. Граф — это фигура, состоящая из точек, называемых вершинами, и отрезков, соединяющих некоторые из вершин, и называемых ветвями. Граф, каждая вершина которого соединена определенной цепью с любой другой вершиной, называется деревом.

К ИПЯ иерархического типа относят традиционные библиотечно-библиографические классификации УДК, ББК, Десятичная классификация Дьюи (США), Международная классификация изобретений (МКИ) и другие. Недостатком иерархических ИПЯ является невозможность организации внеиерархических связей между ЛЕ, однако поисковые возможности могут быть увеличены за счет введения искусственных грамматических средств.

3.2. Неиерархические ИПЯ (или ИПЯ неиерархической структуры) являются языками, в которых сильными парадигматическими отношениями связаны только отдельные «пучки» ЛЕ и есть ЛЕ, не входящие в парадигматику ИПЯ. ЛЕ в таких ИПЯ упорядочивают по внешним признакам, например, в алфавитном порядке. Из содержания документов выбираются ключевые слова, которые преобразуются в ЛЕ данного ИПЯ, затем между ними устанавливаются сильные и слабые парадигматические отношения, которые реализуются в виде ссылочно-справочного аппарата, не охватывающего всех ЛЕ. В качестве примера приводят язык предметных рубрик в дескрипторных ИПЯ и др.

3.3. Фасетные ИПЯ (или ИПЯ фасетной структуры) — это языки, в которых ЛЕ предварительно группируются в фасеты, а иерархические отношения устанавливаются внутри фасетов.

Фасет, (в буквальном смысле означающий — аспект, грань), представляет собой группу однородных терминов, связанных общностью какого-либо признака (основания деления, характеристики). ЛЕ внутри фасетов упорядочиваются иерархически. Фасеты, следующие друг за другом в определенной последовательности, образуют фасетную классификацию.

Большой вклад в развитие теории и практики создания фасетных классификаций внес известный индийский библиотековед Ш. Ранганатан. Наибольшее распространение фасетные классификации получили в Великобритании.

ИПЯ данного типа по сравнению с предыдущими обладают большой семантической силой, но они трудоемки как в разработке, так и в эксплуатации.

Все выше перечисленные классификации ИПЯ изначально были ориентированы на ручной одноаспектный поиск информации. Это вызвало необходимость разработки специальных языков, которые позволяли бы вести многоаспектный поиск с помощью компьютерной техники.

3. ИПЯ дескрипторного типа

3.1. Определение. Методика построения.

Дескрипторные ИПЯ — это искусственные языки, появившиеся в начале 50-х гг. 20 века. В эти годы были разработаны первые механизированные системы поиска информации и предприняты попытки применения ЭВМ для решения разнообразных информационно-поисковых задач. Эти языки появились объективно, т.к. разработанные к тому времени языки не соответствовали требованиям механизированного и автоматизированного поиска информации.

В основу его создания положена научная гипотеза о том, что основным носителем информации в любом тексте являются ключевые слова, под которыми понимаются все члены предложения, несущие основную смысловую нагрузку в тексте. Ключевыми словами могут быть не все, а только некоторые части речи, обозначающие понятия об объектах информации и действиях над ними. В результате того, что не только в естественном языке, но и в ИПЯ отдельные ключевые слова могут быть скоординированы друг с другом в необходимых сочетаниях для получения текстов с заданным смысловым содержанием, было решено создавать такие ИПЯ, в которых лексическими единицами были бы слова, а не рубрики. Такие ИПЯ, называемые еще языками координатного индексирования, получили широкое распространение.

Отметим, что основы ИПЯ такого типа и соответствующие технологии были разработаны в США в 50-е гг. прошлого века учеными М. Таубе и К.Муэрсом[4]. Преимущества данного метода заключаются в следующем: отпадает необходимость в классификационных схемах и перечнях предметных рубрик;индексирование новых документов освобождается от субъективизма — ключевые слова выбираются формально, что можно выполнять и в автоматизированном режиме. К основым достоинствам такого подхода к раскрытию содержания документов и поиску информации заключаются в том, что он позволяет находить информацию по любому, заранее не предвиденному сочетанию

[4] Гиляревский Р.С. Основы информатики: Курс лекций / Р.С. Гиляревский.— М.: Изд-во «Экзамен», 2003. — С.143-145.

признаков. При этом специалисты отмечают и дополнительные трудности, которые сопровождают внедрение дескрипторных ИПЯ. Прежде всего поиск с использованием естественного языка ограничивает его рамками знакомых пользователю языков. Чтобы расширить этот круг, приходится прибегать к словарям. Кроме того, каждый естественный язык отличается богатством своего словарного состава — слова, одинаковые по написанию, могут иметь разный смысл (многозначность, омонимия), а одно и то же понятие может выражаться разными терминами (синонимия).

Термины находятся в сложных взаимоотношениях между собой, выражают более узкие или более широкие понятия, могут быть связвны по сходству, по контрасту или по другим ассоциациям. Для учета этих отношений необходимо составлять на каждом языке специальные понятийные справочники или словари, которые называются тезаурусы. В них для каждого понятия (класса условной эквивалентности) выбирается один термин — дескриптор, а для остальных слов указывается их связь с дескриптором. Тезаурусы еще называют дескрипторными словарями, а сам поиск с их использованием — дескрипторным. Кроме словарей для поиска по ключевым словам и дескрипторам часто создают специальную грамматику, необхобимость в которой обусловлена возникновением ложной кооржинации терминов, ошибочным их сочетанием и т.п.

Подробнее рассмотрим основные понятия дескрипторных ИПЯ.

Понятие дескриптора как термина, предназначенного для однозначного описания понятий, как уже было отмечено, ввел в информатику К. Муэрс. Именно этот ученый предложил учитывать синонимию и применять лексикографический контроль за используемыми ключевыми словами, который заключался в их нормализации и полном устранении неоднозначности и многозначности. Он также считал, что для контроля лексики необходим специальный дескрипторный словарь

Дескрипторный ИПЯ — это искусственный язык координатного типа, построенный на базе формализованной лексики естественного языка.

Для контроля лексики создается информационно-поисковый тезаурус, понимаемый как контролируемый словарь лексических единиц дескрипторного языка, основанный на лексике естественного языка, отображающийсемантические отношения между лексическими единицами и предназначенный для организации поиска информации путем индексирования документов и запросов. Тезаурус — это особый вид идеографических словарей или лингвистических тезаурусов, принципы построения которого были разработаны в лингвистике еще до появления возможности машинной обработки информации. В основе составления словарей такого типа лежит смысловая классификация лексики. Каждый конкретный тезаурус отражает систему понятий определенной области

знаний в виде структурированной совокупности терминов и является результатом большой трудоемкой работы высококвалифицированных специалистов.

Тезаурус предназначен для выполнения следующих функций: индексирования документов и запросов, т.е. перевода их содержания с естественного на дескрипторный язык; фиксирования парадигматических отношений между дескрипторами, что существенно расширяет возможности информационного поиска.

Тезаурусы различаются по различным признакам. В зависимости от тематического профиля выделяют — многоотраслевые, отраслевые, тематические. По своему назначению они бывают рабочими и базовыми. Рабочие используются в реальных системах поиска, обслуживающих определенных пользователей. Базовые являются лексико-семантической основой при построении узкотематических рабочих тезаурусов. В качестве базовых тезаурусов выступают многоотраслевые и отраслевые, включающие основую лексику и основные парадигматические отношения. Рабочие тезаурусы отличаются большой полнотой лексики и более развитой парадигматикой.

Составление тезауруса, как уже было отмечено, это большой трудоемкий процесс, требующий профессиализма и творчества. Однако методика его построения, правила, структура, состав и форма — обеспечены нормативными документами, стандартами, в соответстви с которыми выделены следующие этапы построения тезауруса:

1. определение тематического профиля;
2. сбор лексики и формирование словника ключевых слов;
3. построение словарных статей и формирование лексико-семантических указателей;
4. разработка вспомогательных указателей;
5. оформление тезауруса;
6. его экспертиза и регистрация.

Кратко охарактеризует эти этапы.

1. Тематический профиль тезауруса определяется путем анализа информационных потребностей специалистов той или иной отрасли, обслуживаемых соответствующей службой информации. Разработка нового тезауруса предпринимается только в том случае, если отсутствует тезаурус по заданной тематике или по каким-либо причинам непригоден уже имеющийся аналог. Все разработанные тезаурусы должны регистрироваться в одном месте, так, в советский период их регистрировало подразделение ВНИИКИ (Всесоюзного НИИ технической информации, классификации и кодирования).

2. Второй этап предполагает сбор лексики и формирование словника ключевых слов. Словник понимается как массив терминов, который подвергается семантической обработке в процессе дескрипторизации. Этот массив формируется путем извлечения из

текстов (первичных и вторичных документов) — ключевых слов, т.е. в процессе индексирования документов, а также запросов. Иногда для набора терминов используется справочная литература, таблицы классификации, опрос специалистов и пр. источники.

На этом этапе перед разработчиками возникают две основные проблемы:

- проблема ключевых и неключевых слов;
- проблема формулировки ключевых слов.

Согласно методическим рекомендациям, в качестве ключевых слов могут использоваться имена существительные, прилагательные, причастия, наречия, глаголы (в форме инфинитива). Однако и в стандартах не было выработано единого методического решения о порядке слов в словосочетаниях: прямой или инфертированный (обратный) порядок.

3. Этап построения словарных статей и формулирования лексико-семантического указателя — заключается в дескрипторизации ключевых слов и установлении парадигматических отношений между дескрипторами.

Дескрипторизация осуществляется с целью построения дескрипторного словаря, а установление парадигматических отношений — для увеличения семантической силы языка. В процессе выполнения операции дескрипторизации устраняется неоднозначность ключевых слов (в виде полисемии или синономии). После группировки ключевых слов в классы условной эквивалентности производится выбор одного из включенных терминов в качестве дескриптора. Критериями выбора дескриптора служат полнота выражения смыслового значения данного класса, краткость и понятность, частота встречаемости термина в текстах документа и запроса.

Дескриптор — это лексическая единица тезауруса, под которой принято понимать нормализованное слово или словосочетание, выбранное из множества условно-эквивалентныхключевых слов для его обозначения.

Аскриптор (недескриптор) — лексическая единица тезауруса, входящая в класс эквивалентности данного дескриптора, которая при индексировании документов и запросов подлежит замене на дескриптор.

Результатом дескрипторизации является дескрипторный словарь, который представляет собой алфавитный перечень дескрипторов и аскрипторов.

Процесс установления парадигматических отношений сводится к логическому, ассоциативному и прагматическому анализу дескрипторов. Логический анализ заключается в сопоставлении объемов понятий, представленных дескрипторами, с целью выявления сильных логических отношений: подчинения и пересечения. Эти отношения фиксируются в словарной статье дескриптора с помощью ссылок: выше (в), ниже (н),

выше род (вр), ниже вид (нв). Ассоциативный анализ заключается в сопоставлении признаков предметов, входящих в определенные понятия, представленные дескрипторами, с целью выявления слабых ассоциативных отношений. Эти отношения записываются в словарной статье дескриптора с помощью ссылки ассоциация (а). Прагматический анализ осуществляется с целью упорядочения состава и структуры словарной статьи дескриптора, четкого разграничения сильных и слабых парадигм.

В результате третьего этапа разработки тезауруса формируется лексико-семантический указатель, который представляет собой упорядоченую в алфавитном порядке последовательность словарных статей. Дескрипторная словарная статья сосотоит из заглавного дескриптора, синонимичных ему дескрипторов и аскрипторов, связанных с ним парадигматическими отношениями.

Структура дескрипторной статьи определяется формулой:

Д {М с; М в; М н; М а}

где Д — заглавный дескриптор;

М с — множество аскрипторов (синонимов), входящих в класс эквивалентности дескриптора;

М в — множество вышестоящих дескрипторов;

М н — множество нижестоящих дескрипторов;

М а — множество ассоциадескрипторов.

Внутри каждой группы лексических единиц, связанных с заглавным дескриптором одним из видов связи, устанавливается алфавитный порядок расположения. Например, словарная статья дескриптора — БИБЛИОТЕКИ.

БИБЛИОТЕКИ

с библиотечные службы

библиотечные учреждения

библиотечные центры

в Информационные службы

н Массовые библиотеки

Научные библиотеки

Специальные библиотеки

а Библиотечно-библиографическое обслуживание

Словарная статья аскриптора состоит из аскриптора и заменяющих его и их комбинаций. Например,

библиотечные службы

см. БИБЛИОТЕКИ

Дескрипторные и аскрипторные статьи располагв лексико-семантической части тезауруса в алфавитном порядке.

4. Этап — составление вспомогательных указателей, среди которых наиболее распространенными считаются систематический указатель дескрипторов тезауруса (совокупность алфавитных списков дескрипторов); указатель иерархических отношений (свод классификационных иерархических деревьев, считается удобным средством контроля при пополнении тезауруса); пермутационный указатель дескрипторов (предназначенный для поиска лексической единицы по отдельным словам).

5. Этап — оформление тезауруса, которое выполняется в соответствии с существующим национальным стандартом и с учетом требований международных стандартов.

6. Этап — экспертиза и регистрация тезауруса.

3.2. Грамматика дескрипторных ИПЯ

Дескрипторные ИПЯ являются, как уже было сказано, языками координатного типа и способны обеспечить многоаспектный информационный поиск. Однако, следует отметить, что при поиске возможны ложные сочетания дескрипторов внутри поиского образа документа, т.к. они не связаны никакими текстуальными отношениями. Ложные сочетания искажают смысл документа и вызывают информационный шум в системе, т.е. выдачу нерелевантных документов.

Основным способом уменьшения информационного шума является введение в ИПЯ грамматических средств, которые позволяют выражать синтагматические отношения.

Все известные грамматические средства дескрипторных ИПЯ можно подразделить на :

- *фрагментирующие* — для разделения поискового образа документа на части; и

- *смыслоразличительные* — для указания смысловой роли различительных слов внутри фрагмента поискового образа документа.

К основному фрагментирующему средству относят *указатели связи*, представляющие собой символ (букву, цифру, знак пунктуации), который присоединяется к дескрипторам поискового образа документа, входящим в один фрагмент. Документ подразделяется на нексколько тем и набор дескрипторов каждой темы получает свой указатель связи. Некоторые дескрипторы могут входить в разные фрагменты, поэтому они получают несколько указателей связи. При поиске указатели связи дают возможность группировать дескрипторы, относящиеся к одной теме.

К смыслоразличительным грамматическим средствам относят: “мешочную” грамматику и позиционную грамматику.

“Мешочная” грамматика — это простое перечисление дескрипторов в поисковом образе документа и сам факт этого перечисления свидетельствует о том, что между ними есть связь, но она не выражается. Такая грамматика используется в узкотематических и отраслевых дескрипторных ИПЯ технической тематики. Это практически язык без грамматики.

Позиционная грамматика. Ее суть заключается в том, что поисковый образ документа строится с помощью специальных формул, матриц и представляет собой кортеж (цепочку, линейную запись) дескрипторов, месторасположение которых регламентировано определенными правилами. Цепочек может быть столько, сколько в документе отдельных тем. Синтагматические отношения между дескрипторами выражаются через их последовательность в кортеже. Например, первый дескриптор в кортеже может обозначать вид действия второй — место действия, третий — время действия и т.д. Очевидно, что в кортеже каждый элемент может иметь только один предшествующий и один последующий дескриптор. Поэтому такой метод не позволяет выражать сложных синтагматических отношений между дескрипторами поискового образа документа.

Одним из важнейших смыслоразличитльных грамматических средств является *указатель роли,* который позволяет опрделить, какое значение имеет тот или иной дескриптор в поисковом образе документа. Обозначаются указатели роли каким-либо символом (буквой, цифрой или их комбинацией), которые приписываются дескрипторам или ключевым словам в поисковом образе документа в зависимости от грамматической категории, к которой относится тот или иной дескриптор. Основными грамматическими категориями могут быть “процесс”, “свойства”, “материя”, “среда” и пр.

По мнению специалистов, введение в дескрипторый ИПЯ грамматических средств значительно усложняет процесс индексирования и информационного поиска, делает их более дорогостоящими. Считается, что для небольших информационно-поисковых массивов выгоднее применение дескрипторных ИПЯ без грамматики. Такие выводы были сделаны в период преобладания “ручного” поиска и ИПЯ.

Современный уровень развития информатизации общества обусловил переход на автоматизированный поиск в ИПС: локальных или сетевых. В таких условиях напротив — возростает значение грамматических средств в дескрипторных и других ИПЯ, которые используются разработчиками ИПС, в том числе и для сети Интернет, т.к. их присутствие существенно повышает эффективность информационного поиска и снижает информационный шум.

4. Лингвистическое обеспечение информационных систем

Лингвистическое обеспечение (или средства) — это одно из основых средств обеспечения любой автоматизированной информационой системы (АИС) наряду с техническим, информационным, программным обеспечением. Оно включает: язык описания данных (ЯОД), язык манипуляции данных (ЯМД), язык запросов, средства и методы индексирования, словари.

Пользователи АИС применяют языковые средства, которые делятся на две группы:

- алгоритмические языки запросов;
- дескрипторно-кодовые языки запросов.

Такое деление обусловлено разделением пользователей на специалиство и неспециалистов в области информатики и программирования.

Алгоритмические языки запросов в свою очередь также поделены на две группы: ЯОД и ЯМД.

ЯОД — языки высокого уровня, предназначенные для задания схемы базы данных (БД). С их помощью описываются типы данных, подлежащих хранению в БД, их структура и связи.

ЯМД — языки запросов к БД командного типа, определяющие операции, выполняемые над данными. Они представляют собой подмножество русского языка.

Дескрипторно-кодовые языки представлена языками классификационного и дескрипторного типов. В основу классификационных языков положены систематические классификации, отражающие смысловые отношения между понятиями. Примерами могут служить известные библиотечные классификации УДК, ББК и другие. Они считаются эффективными при широкотематическом поиске. Идея дескрипторных ИПЯ, как было уже сказано выше, состоит в том, что содержание документа выражается ограниченным набором слов, встречающихся в тексте, если эти слова наиболее точно выражают его индивидуальные особенности. Другими словами, в основе дескрипторного языка лежит алфавитный перечень ключевых слов.

По типу лексического состава ИПЯ делят на языки с фиксированными и свободноразвивающимися словарями. В первом случае тезаурус содержит ограниченный список дескрипторов. Во втором случае в ПОД включаются и новые термины.

Следующей составляющей лингвистического обеспечения АИС являются методы и средства индексирования. Индексирование понимается как присвоение документу ключевых слов или кодов, служащих указателями документов и используемых для его поиска. Более широкое понимание индексирования включает сжатое изложение текста на естественном языке.

К методам индексирования относят:

1. Извлечение — компьютерный анализ лексического состава текста, выбор из него слов и словосочетаний, удовлетворяющих заранее установленным критериям.

2.Приписное индексирование — сравнение лексического состава текста с индексными терминами класиификационной схемы и приписка одной или нескольких предметных рубрик.

3.Автоиндексирование. К его средствам относят средства, использующие информацию, извлекаемую из естественного текста и дополнительные сведения из словатей, составленных на основе обработки больших массивов политематической лексики языка.

Данные лексические средства приводят к формализованному виду, устанавливают связи между элементами.

Любая АИС должна быть обеспечена возможностью использования синонимов и других парадигматических отношений между терминами, т.е. всех тех средств, которые обычно представляются тезаурусом. Применение тезауруса позволяет в значительной мере смягчить требования к формулировке запросов с точки зрения обеспечения поиска необходимых документов в условиях использования синонимичных терминов для обозначения одних и тех же объектов. Над индексами, присвоенными документу, возможны логические операции с помощью операторов «И», «ИЛИ», «НЕ» (дизъюнкции, конъюнкции и отрицания).

К лексическим средствам АИС относятся также словари, являющиеся вспомогательными средствами для обработки информации на компьютере, упрощающих данную процедуру. Они бывают различных видов: словарь основ (или нулевой словарь), который используется для автоматического контроля лексики естественного языка; словарь синонимов; словарь префиксов; словарь запретов, устраняющий предлоги, союзы местоимения, наречия, служебные слова.

Лингвистические средства позволяют пользователю описать постановку задачи на ИПЯ, корректировать ее, повторять отдельные этапы работ, организовывать решение альтернативными путями.