

2-я лекция.

Типология, структура и функция информационных систем

На первой лекции мы рассмотрели значение термина «информационная система» и выяснили, что его можно понимать по-разному. Во-первых, каждой «системе» свойственны информационные процессы, и поэтому каждая система – «информационная». Большой класс систем используют информацию как основное **средство** функционирования. К ним относятся, собственно, все социальные системы. Среди технических систем таковы «автоматизированные системы управления (социальными объектами) – АСУ» и «системы автоматического управления (техническими объектами) – САУ», поскольку «управление» и есть выработка управляющей информации. Но имеются и системы, у которых информация является не только свойством, не только средством, но **целью** функционирования – «информационные системы» в чистом виде.

Можно предположить, что среди социальных систем есть системы, способные **рождать** информацию, у которых информация становится **продуктом** деятельности. Такое предположение правдоподобно для систем искусства (в частности – литературы), религии, систем власти... Человеческое **мышление** тоже вроде бы способно рождать информацию. В искусственных системах феномен рождения информации, увеличение количества новой информации в процессах её обработки внутри системы никогда не наблюдается. В лучшем случае можно добиться **компрессии** информации, т. е. выделения существенной и удаления ненужной информации, но при этом общее количество информации не возрастает, а именно **убывает**. Можно сформулировать **общий закон убывания информации** в разных вариантах:

- количество информации, поступающей на входы системы, больше количества информации на выходах системы
- количество информации в замкнутой системе (не имеющей входов) в процессах обработки уменьшается (не возрастает)
- информация не может возникнуть «из ничего».

Теоретически этот закон следует из нашего представления об информации как о **форме материи, передающейся в процессах взаимодействия**. Поскольку «форма» это то же самое что «разнообразие частей», то сформулированный выше закон можно изложить в общефизических терминах без понятия информации:

- в замкнутой системе разнообразие частей со временем уменьшается.

А это и есть известный в физике «закон возрастания энтропии¹». Его доказательство при некоторых естественных предположениях может быть получено строго математически. В научной информатике математическая строгость вывода имеет второстепенное значение, а закон убывания информации (возрастания энтропии) может быть принят как эмпирический (опытный) факт.

Но тогда мы должны **искать источник** информации в тех случаях, когда нам представляется её рождение «из ничего». И **религия** не только пользуется информацией из этого источника, но и называет его – **Бог**. Другие (философские, социологические, биологические) обоснования творческого начала информационных процессов в других терминах также вводят некоторую трансцендентную (потустороннюю) силу, не подлежащую научному осмыслению. В курсе научной информатики мы не будем исследовать то, что не подлежит научному осмыслению. Следовательно наш объект изучения – **системы не создания, а переработки, передачи информации**.

Здесь однако следует разобраться в особенностях работы с информацией в разных системах. Согласно нашему определению информации как **формы в процессе её передачи** информация (каждая её порция) связана как с объектом-источником, так и с объектом-приёмником передаваемой формы. Т. е. каждая порция информации в собственном смысле слова есть сложный объект, который трудно охватить в целостности, включая и источник, и приёмник, и форму. Оторвав форму от источника и приёмника, получаем более простой объект, называемый **данными**. С таким объектом работать проще. **Системы обработки данных** – в частности **компьютеры и сети связи** – являются основой информационной технологии, но только именно основой, а не вершиной, не смыслом информационной деятельности. Изучение и применение систем и процессов **обработки данных** – это большая и важная часть современных технологий. Но мы будем изучать системы, которые реализуют информационные процессы **с помощью** в частности и систем обработки данных.

Таким образом, мы должны представлять истинно информационные системы в окружении с одной стороны **источников информации**, а с другой стороны – **потребителей информации**, включая тех и других в состав объекта изучения. Источниками информации являются **документы** – материальные объекты с записанной на них информацией. Потребителями могут быть различные коллективы или категории людей, а также организации и даже некоторые автоматизированные системы. В зависимости от характера окружения (анализируемых документов и обслуживаемых потребителей) различается большое количество типов информационных систем, обладающих разными свойствами и внешне несопоставимых друг с другом, но внутренне сходных по составу функций и структуре элементов.

¹ Энтропия – это принятая в физике численная характеристика однородности (отсутствия разнообразия частей) вещества.

Если система ставит перед собой задачу обработки всего универсума документов и обслуживает весь универсум пользователей, то это представляется как **универсальная публичная библиотека**. Правда, библиотеки обычно имеют дело только с письменными документами, но в последнее время включают также и аудиовизуальные. Системы, работающие с нетекстовыми документами – это **музеи**.

Системы, работающие с документами в Интернете и обслуживающие посетителей Интернета, – это известные поисковые машины (Яндекс, Гугл и др.), а также специализированные порталы, предоставляющие доступ к ресурсам определённого вида или тематики.

Системы, ограничивающие свой вход научной и технической документацией и обрабатывающие её с помощью вычислительной техники, составляют класс автоматизированных систем научно-технической информации – **АСНТИ**.

Системы, ограничивающие свой вход распорядительной документацией и производственно-хозяйственными данными, называются автоматизированными системами организационного управления (АСУ), хотя в их состав стараются включать также функции не только поиска информации, но также функции выработки управляющих решений. Управляющие функции АСУ обычно сводятся к поиску и предоставлению заготовок (форм, бланков) распорядительных документов, которые должны быть заполнены лицом, находящимся вне системы. Так что АСУ также являются по преимуществу системами не создания управляющей информации, а системами поиска и передачи пользователям информации, необходимой для управления.

В нашей стране в советский период, когда руководство экономикой было строго централизовано и распределено по отраслям, возглавляемым министерствами, при каждом министерстве создавались **отраслевые АСУ** и АСНТИ, работавшие с документами определённой хозяйственной отрасли и обслуживавшие предприятия этой отрасли.

Мировую известность и мировое распространение приобрели некоторые зарубежные АСНТИ, работающие с документами определённой **тематики**, такие как INIS (ядерные исследования), AGRIS (сельское хозяйство), MEDLARS (медицина) и др.

На крупных предприятиях, фирмах и в исследовательских центрах создаются корпоративные информационные системы.

Этот фрагментарный обзор показывает разнообразие информационных систем, подлежащих нашему рассмотрению. Во всех этих системах ядром технологии является некоторый программно-технологический комплекс, выполняющий автоматизированные информационные процедуры с помощью вычислительной техники. Но кроме вычислительной техники во все эти системы входят и другие компоненты, в частности люди, без которых работа систем была бы невозможна. Неавтоматизированные компоненты являются

существенными даже для таких сугубо автоматических систем как поисковые машины Интернета. Внутреннюю структуру информационной системы мы в общих чертах рассматривали на прошлой лекции. Вот эта схема:

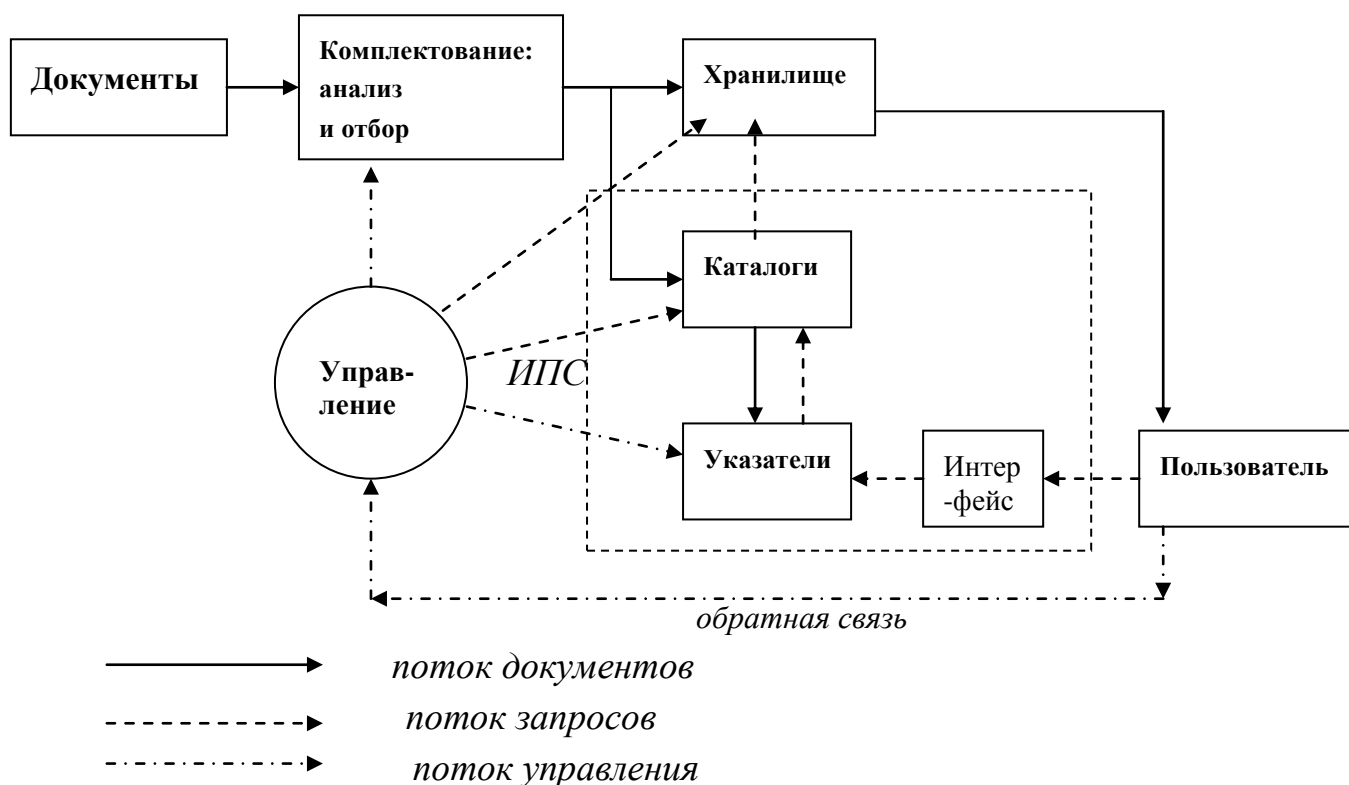


Рис. 2.1. Документальная информационная система

В схеме следует предусмотреть также управляющую подсистему, связанную с другими компонентами управляющими воздействиями и получающую от пользователей информацию с оценкой действия системы (так называемая **обратная связь**).

Каждый компонент, изображённый на схеме представляет собой в свою очередь сложный объект – подсистему в рамках общей системы. Давайте их рассмотрим по-подробнее.

Множество документов включает разные виды документов, обладающие различными свойствами и действующими в разных областях жизни. Типичная информационная система принимает документы различных видов, поэтому блок комплектования естественно разбивается на подблоки в зависимости от вида обрабатываемых документов. Даже в тех случаях, когда система ограничивается каким-либо одним видом документов (например, международная система патентной документации IMPADOC), на поверку выходит, что в неё поступают разные виды документов. В данном случае, например, патентные документы внутри себя делятся на (1) патентные заявки, (2) описания зарегистрированных изобретений, (3) информационные сообщения о заявках и изобретениях. Сюда также входят заявки и описания товарных знаков, полезных моделей и других объектов интеллектуальной

промышленной собственности. Каждый из видов документов имеет свои особенности выявления, сбора и обработки. Поэтому сам блок комплектования информационной системы представляет собой обычно совокупность подблоков (подсистем), которые ведут работу со своим видом документов, и технологические процессы в которых могут сильно различаться. Для нас важно, что современные информационные системы обрабатывают как традиционные документы, так и электронные документы. Особенно специфична работа с сетевыми документами (сайтами Интернета).

Поисковые машины Интернета являют собой чистый образец работы с сетевыми ресурсами. Подсистема комплектования в них реализована как робот («паук»), который периодически обегает заданный ему фрагмент сети или всю сеть, прочитывает доступные ему части сайтов, составляет описание каждого сайта в виде совокупности обнаруженных в нём слов и направляет это описание в блок каталога вместе с сетевыми адресами страниц, на которых данные слова найдены. Особенность сетевых поисковых машин состоит в том, что они обычно не формируют своего хранилища, а используют в качестве блока хранения всю сеть целиком. Но здесь мы видим наличие ещё одной подсистемы, которая на общей схеме рисунка 2.1 не была изображена – блок доставки документов пользователю. Очевидно, что такая подсистема должна существовать в случае материальной почтовой доставки книжных изданий. Для поисковых машин было бы достаточно обойтись общесетевыми средствами доступа к источнику по известному адресу. Однако многие поисковики предлагают предварительную обработку документа в двух видах: (1) дают минимальное текстовое окружение найденных слов и (2) предоставляют текст документа с размеченными вхождениями слов запроса, что позволяет сразу обращаться к интересующему пользователя фрагменту без необходимости просматривать весь документ целиком. Схему сетевой поисковой машины можно представить рисунком 2.2.

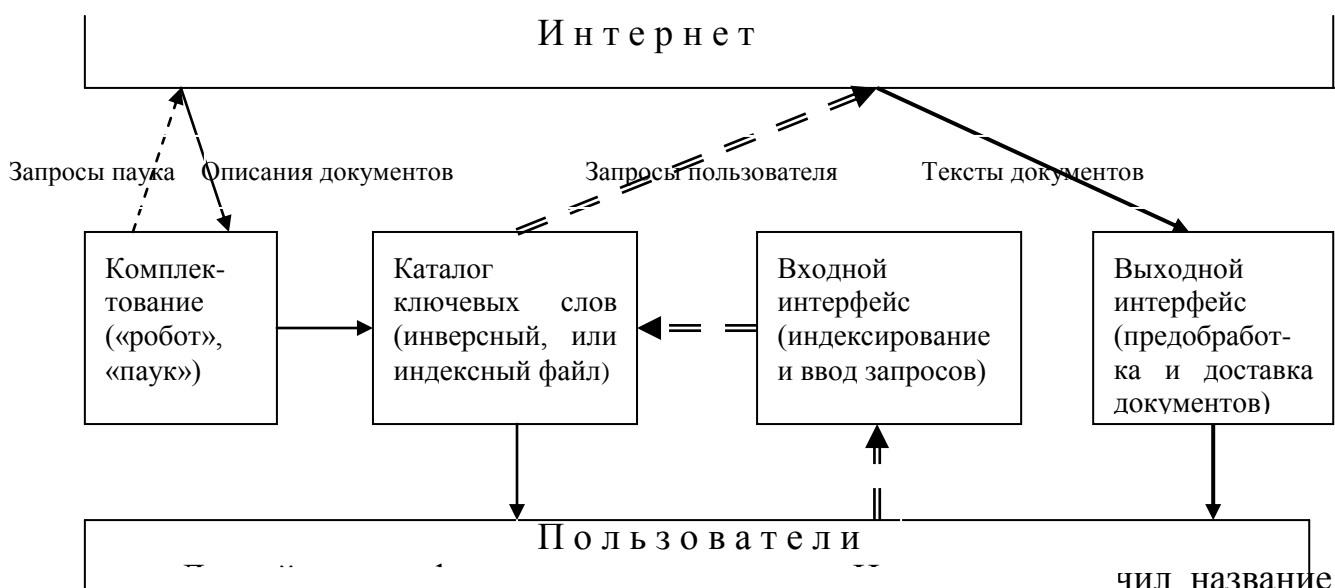


Рис. 2.2. Структура и функции поисковой машины

чил название
порталов от

поисковых машин состоит в том, что порталы открывают доступ к документам определённой смысловой области. Соответственно этому каталог порталов построен не на формальном выявлении слов в документах, а на **отборе** документов по определённым семантическим категориям. Если каталог поисковой машины моделирует предметный каталог традиционной библиотеки, то каталог сетевого портала моделирует систематический каталог библиотеки. Соответственно входной интерфейс порталов должен пользоваться указателем семантических классов каталога, т. е. тематической классификационной системой. Каждый отображенный в портал документ должен получить тематический индекс по данной классификационной системе. Вопрос об автоматической классификации в настоящее время не решён, и видимо, не будет в полной мере решён никогда. Поэтому в порталах может быть организован доступ только к ограниченному массиву документов, который удастся просмотреть и оценить силами экспертов, составляющих персонал портала. В противоположность этому поисковые машины включают в свой доступ практически весь объём Интернета в течение примерно недели.

Ведущие поисковые машины стараются включить в себя также функции семантического отбора, свойственные порталам. Для этого наряду с пословным каталогом в них организуется и тематический каталог, составленный из наиболее популярных тем поиска в Интернете (например, «музыка», «отдых», «образование», «здоровье», «наука», ...) но в этих каталогах также отражается только незначительная часть сетевых информационных ресурсов.