

Святослав Милованов

# **Информационные системы. Введение в дисциплину**

Святослав Милованов

**Информационные системы.  
Введение в дисциплину**

«Автор»

2026

**Милованов С. С.**

Информационные системы. Введение в дисциплину /  
С. С. Милованов — «Автор», 2026

Курс лекций разработан на кафедре прикладной информатики и информационного менеджмента Поволжской академии государственной службы им. П.А. Столыпина для студентов специальности 350800 «Документоведение в управлении». Соответствует ГОС. Информация содержащаяся в данном учебно-методическом пособии будет полезна также и для студентов других специальностей.

© Милованов С. С., 2026

© Автор, 2026

# Содержание

|                                   |    |
|-----------------------------------|----|
| Глава                             | 5  |
| Конец ознакомительного фрагмента. | 21 |

# Святослав Милованов

## Информационные системы.

### Введение в дисциплину

#### Глава

|                                                                                   |                                                                  |                                                                 |                                                        |                                                   |              |           |                                                                                                                                 |                                                                                                                                                                                                                                             |                                                                                                                         |                                                                                                                                                                                                                                                                                                                                                              |                                                                              |                    |
|-----------------------------------------------------------------------------------|------------------------------------------------------------------|-----------------------------------------------------------------|--------------------------------------------------------|---------------------------------------------------|--------------|-----------|---------------------------------------------------------------------------------------------------------------------------------|---------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|-------------------------------------------------------------------------------------------------------------------------|--------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|------------------------------------------------------------------------------|--------------------|
| РОССИЙСКАЯ АКАДЕМИЯ ГОСУДАРСТВЕННОЙ СЛУЖБЫ ПРИ ПРЕЗИДЕНТЕ<br>РОССИЙСКОЙ ФЕДЕРАЦИИ | ПОВОЛЖСКАЯ АКАДЕМИЯ ГОСУДАРСТВЕННОЙ<br>СЛУЖБЫ им. П.А. Столыпина | Кафедра прикладной информатики и информационного<br>менеджмента | Информационные системы. Введение в дисциплину.         | учебно-методические<br>рекомендации для студентов | Саратов 2004 | УДК ББК Д | Информационные системы. Введение<br>в дисциплину. Учебно-методические рекомендации для студентов. – Саратов: ПАГС, 2004. – 99с. | Курс лекций разработан на кафедре прикладной информатики и информационного менеджмента Поволжской академии государственной службы им. П.А. Столыпина для студентов специальности 350800 «Документоведение в управлении». Соответствует ГОС. | Информация содержащаяся в данном учебно-методическом пособии будет полезна также и для студентов других специальностей. | Составители: кандидат медицинских наук, доцент Милованов С.С. Рецензенты: канд. физ.-мат. наук, доцент кафедры математической физики и вычислительной математики СГУ С.Н. Кабанов; канд. пед. наук, доцент кафедры математики и статистики ПАГС Ю.Б. Соколовская. Утверждены и рекомендованы к изданию Центральным координационно-методическим советом ПАГСа | Составители: Милованов С.С. Поволжская академия государственной службы, 2004 | Содержание. стр.1. |
| <b>Информация,</b>                                                                | <b>свойства</b>                                                  | <b>информации.</b>                                              | .....                                                  | 51.1                                              |              |           |                                                                                                                                 |                                                                                                                                                                                                                                             |                                                                                                                         |                                                                                                                                                                                                                                                                                                                                                              |                                                                              |                    |
| История                                                                           | и                                                                | теории                                                          | информации.                                            | .....                                             | 51.2         |           |                                                                                                                                 |                                                                                                                                                                                                                                             |                                                                                                                         |                                                                                                                                                                                                                                                                                                                                                              |                                                                              |                    |
| Применение                                                                        | теории                                                           | информации.                                                     | .....                                                  | 51.3                                              | Схема        |           |                                                                                                                                 |                                                                                                                                                                                                                                             |                                                                                                                         |                                                                                                                                                                                                                                                                                                                                                              |                                                                              |                    |
| передачи                                                                          | информации.                                                      | .....                                                           | 61.4                                                   | Кодирование                                       |              |           |                                                                                                                                 |                                                                                                                                                                                                                                             |                                                                                                                         |                                                                                                                                                                                                                                                                                                                                                              |                                                                              |                    |
| информации.                                                                       | .....                                                            | 71.5                                                            | Основные определения                                   |                                                   |              |           |                                                                                                                                 |                                                                                                                                                                                                                                             |                                                                                                                         |                                                                                                                                                                                                                                                                                                                                                              |                                                                              |                    |
| темы                                                                              | .....                                                            | 82.                                                             | Документальные информационно-поисковые системы (ДИПС). | .....                                             | 112.1        | Теория    | ДИПС.                                                                                                                           |                                                                                                                                                                                                                                             |                                                                                                                         |                                                                                                                                                                                                                                                                                                                                                              |                                                                              |                    |
| .....                                                                             | .....                                                            | 112.2                                                           | Общая функциональная                                   |                                                   |              |           |                                                                                                                                 |                                                                                                                                                                                                                                             |                                                                                                                         |                                                                                                                                                                                                                                                                                                                                                              |                                                                              |                    |
| структура                                                                         | ДИПС.                                                            | .....                                                           | 122.3                                                  | Представление                                     |              |           |                                                                                                                                 |                                                                                                                                                                                                                                             |                                                                                                                         |                                                                                                                                                                                                                                                                                                                                                              |                                                                              |                    |
| содержания                                                                        | текста.                                                          | .....                                                           | 122.4                                                  | Информационно-поисковые языки (ИПЯ)               |              |           |                                                                                                                                 |                                                                                                                                                                                                                                             |                                                                                                                         |                                                                                                                                                                                                                                                                                                                                                              |                                                                              |                    |
| .....                                                                             | .....                                                            | 132.5                                                           | Обработка                                              |                                                   |              |           |                                                                                                                                 |                                                                                                                                                                                                                                             |                                                                                                                         |                                                                                                                                                                                                                                                                                                                                                              |                                                                              |                    |
| информации.                                                                       | .....                                                            | 132.6                                                           | Основные                                               |                                                   |              |           |                                                                                                                                 |                                                                                                                                                                                                                                             |                                                                                                                         |                                                                                                                                                                                                                                                                                                                                                              |                                                                              |                    |
| темы.                                                                             | .....                                                            | 143.                                                            |                                                        |                                                   |              |           |                                                                                                                                 |                                                                                                                                                                                                                                             |                                                                                                                         |                                                                                                                                                                                                                                                                                                                                                              |                                                                              |                    |
| <b>ДИПС.</b>                                                                      | .....                                                            | 153.1                                                           |                                                        |                                                   |              |           |                                                                                                                                 |                                                                                                                                                                                                                                             |                                                                                                                         |                                                                                                                                                                                                                                                                                                                                                              |                                                                              |                    |
| Автоматическое                                                                    | индексирование.                                                  | .....                                                           | 153.2                                                  |                                                   |              |           |                                                                                                                                 |                                                                                                                                                                                                                                             |                                                                                                                         |                                                                                                                                                                                                                                                                                                                                                              |                                                                              |                    |
| Автоматическое                                                                    | рубрицирование.                                                  | .....                                                           | 163.3                                                  | Поиск                                             |              |           |                                                                                                                                 |                                                                                                                                                                                                                                             |                                                                                                                         |                                                                                                                                                                                                                                                                                                                                                              |                                                                              |                    |
| текстовой                                                                         | информации.                                                      | .....                                                           | 203.4                                                  | Методы ведения                                    |              |           |                                                                                                                                 |                                                                                                                                                                                                                                             |                                                                                                                         |                                                                                                                                                                                                                                                                                                                                                              |                                                                              |                    |
| обратной                                                                          | связи                                                            | с                                                               | пользователем.                                         | .....                                             | 213.5        | Оценка    | качества                                                                                                                        |                                                                                                                                                                                                                                             |                                                                                                                         |                                                                                                                                                                                                                                                                                                                                                              |                                                                              |                    |
| ДИПС.                                                                             | .....                                                            | 213.6                                                           | Основные определения                                   |                                                   |              |           |                                                                                                                                 |                                                                                                                                                                                                                                             |                                                                                                                         |                                                                                                                                                                                                                                                                                                                                                              |                                                                              |                    |
| темы.                                                                             | .....                                                            | 224.                                                            | <b>Характеристика</b>                                  |                                                   |              |           |                                                                                                                                 |                                                                                                                                                                                                                                             |                                                                                                                         |                                                                                                                                                                                                                                                                                                                                                              |                                                                              |                    |
| <b>информационных</b>                                                             | <b>ресурсов.</b>                                                 | .....                                                           | 254.1                                                  | Сети                                              |              |           |                                                                                                                                 |                                                                                                                                                                                                                                             |                                                                                                                         |                                                                                                                                                                                                                                                                                                                                                              |                                                                              |                    |
| информационное                                                                    | обслуживание.                                                    | .....                                                           | 254.2                                                  | Глобальная                                        |              |           |                                                                                                                                 |                                                                                                                                                                                                                                             |                                                                                                                         |                                                                                                                                                                                                                                                                                                                                                              |                                                                              |                    |
| Internet.                                                                         | .....                                                            | 264.3                                                           | Сервис                                                 |                                                   |              |           |                                                                                                                                 |                                                                                                                                                                                                                                             |                                                                                                                         |                                                                                                                                                                                                                                                                                                                                                              |                                                                              |                    |
| Internet.                                                                         | .....                                                            | 284.4                                                           | Телеконференции.                                       |                                                   |              |           |                                                                                                                                 |                                                                                                                                                                                                                                             |                                                                                                                         |                                                                                                                                                                                                                                                                                                                                                              |                                                                              |                    |

|                                                                 |        |                                                                                     |
|-----------------------------------------------------------------|--------|-------------------------------------------------------------------------------------|
| FTP. Real Chat. GOPHER.....                                     | 304.5  | Всемирная                                                                           |
| паутина (World Wide Web или WWW).....                           | 334.6  | Hyper-                                                                              |
| G. URL. "Навигация" в сети Web .....                            | 345.   |                                                                                     |
| <b>Информационные сети, принципы построения.....</b>            |        |                                                                                     |
| 375.1 Компьютерные сети.....                                    | 375.2  |                                                                                     |
| Топология компьютерной сетей.....                               | 395.3  | Типы                                                                                |
| построения сетей по протоколам информации.....                  | 435.4  |                                                                                     |
| Пример построения сети.....                                     | 446.   | <b>Виды</b>                                                                         |
| <b>документов, организация хранения и средства доступа.....</b> | 476.1  |                                                                                     |
| Технологии баз данных.....                                      | 476.2  | Технология                                                                          |
| преобразования бумажного документа в электронный.....           | 486.3  | Технологии                                                                          |
| создания каталогов системы Windows.....                         | 486.4  | Технологии,                                                                         |
| обеспечивающие безопасность документов.....                     | 506.5  | Технологии сжатия                                                                   |
| информации.....                                                 | 516.6  | Архивное                                                                            |
| хранение.....                                                   | 517.   | <b>Мировые</b>                                                                      |
| <b>информационные ресурсы.....</b>                              | 547.1  | Средства                                                                            |
| доступа.....                                                    | 547.2  | Кодировки символов                                                                  |
| в Internet.....                                                 | 608.   | <b>Поиск электронных</b>                                                            |
| <b>документов в Internet.....</b>                               | 628.1  | Поиск информации                                                                    |
| в WWW.....                                                      | 628.2  | Поисковые указатели                                                                 |
| (индексы).....                                                  | 638.3  | Сбор информации поисковыми                                                          |
| роботами.....                                                   | 648.4  | Сравнение поисковых каталогов и                                                     |
| указателей.....                                                 | 668.5  | Качество                                                                            |
| поиска.....                                                     | 678.6  | Порталы, рейтинговые                                                                |
| службы, основные поисковые системы.....                         | 688.7  | Рекомендации по                                                                     |
| поиску.....                                                     | 739.   | <b>Электронные издательства</b>                                                     |
| <b>и электронные библиотеки (ЭБ).....</b>                       | 759.1  | Основные направления                                                                |
| и задачи в области создания ЭБ.....                             | 759.2  | Зарубежный опыт                                                                     |
| формирования и предоставления ЭБ.....                           | 779.3  | «Национальная ЭБ»                                                                   |
| - единая программа создания ЭБ.....                             | 789.4  | Правовые вопросы                                                                    |
| при создании ЭБ.....                                            | 809.5  | Экономические вопросы                                                               |
| при создании ЭБ.....                                            | 809.6  | Технические вопросы при                                                             |
| создании ЭБ.....                                                | 8110.  | <b>Коллекция и хранилище</b>                                                        |
| <b>электронных документов.....</b>                              | 8310.1 | WEB – серверы как                                                                   |
| хранилища данных в Internet.....                                | 8310.2 | Размещение информации                                                               |
| на WEB – сервере.....                                           | 8310.3 | База данных WAIS                                                                    |
| в Internete.....                                                | 8410.4 | Работа с информацией                                                                |
| в хранилищах электронных документов.....                        | 8511.  | <b>Протоколы HTTP и</b>                                                             |
| <b>Z39.50.....</b>                                              | 8711.1 | Краткая характеристика                                                              |
| Z39.50.....                                                     | 8811.2 | Hypertext transfer protocol (HTTP)                                                  |
| .....                                                           | 8911.3 | Основные различия между Z39.50 и                                                    |
| http.....                                                       | 8911.4 | Управление доступом на основе HTTP                                                  |
| и Z39.50.....                                                   | 9111.5 | Компоненты ЭБ, управляющие доступом к                                               |
| ресурсам.....                                                   | 9311.6 | Совместное использование протоколов Z39.50 и                                        |
| HTTP.....                                                       | 94     | <b>Список рекомендуемой литературы.....</b>                                         |
| .....                                                           | 96     | <b>1. ИНФОРМАЦИЯ, СВОЙСТВА ИНФОРМАЦИИ.1.1 История и</b>                             |
| <b>теории информации</b>                                        |        | Первоначальное понимание информации как сведений сохранялось                        |
|                                                                 |        | вплоть до середины 20 века. В связи с прогрессом технических средств массовых и     |
|                                                                 |        | других коммуникаций (телефон, телеграф, радио, телевидение и т. п.) и в особенности |

с ростом объёма передаваемых сообщений появилась необходимость их измерения для улучшения условий передачи. Первые попытки измерения количества информации относятся к 20-м гг. XX в. Математическая теория информации была создана в 1948 году, когда К. Шеннон и У. Уивер опубликовали статью «Математическая теория связи» (1949), в которой использовались вероятностные методы для измерения количества информации и были предложены абстрактная схема связи, состоящая из шести компонентов (источника информации, передатчика, линии связи, приёмника, адресата и источника помех), а также теоремы о пропускной способности, помехоустойчивости, кодировании. В вероятностной теории под информацией понимались не любые сообщения, которыми обмениваются люди или передают их по техническим каналам связи, а лишь такие, которые уменьшают неопределённость у получателя. Информационная неопределённость существует тогда, когда из-за неполноты информации необходим выбор одной из двух или большего числа возможностей. Такие процессы имеют место не только в коммуникациях, но и в управлении, познании. Наряду с шенноновским, вероятностным, вариантом математической теории информации появились и другие её варианты – топологический, комбинаторный, динамический, алгоритмический, из которых в приложениях используются лишь комбинаторный и вероятностный. Эти математические подходы к измерению информации описывают лишь знаковую структуру сообщений и поэтому могут быть охарактеризованы как синтаксические теории. Содержательный (смысл, значение) и аксиологический (ценность, полезность) аспекты информации исследуются в семантической и прагматической теориях. Информация, кроме количества, ценности и содержания обладает и другими свойствами (правдивость, достоверность, полнота, глубина, точность, убедительность, доказательность, новизна, эффективность, оптимальность, оперативность, надёжность, выразительность).

**1.2 Применение теории информации.** Теория информации является основой кибернетики и теории управления. В 1948 году Винер предложил информационное видение кибернетики, как науки об управлении и связи в живых организмах, обществе и машинах. В 60-70-х гг. прошлого века появились работы, где не только на содержательном, но и формально-математическом уровне наметилась идея синтеза знаний о связи и управлении в так называемой «информационной теории управления». Понятие информации стало общенаучным понятием, то есть общим для всех частных наук, а информационный подход, включающий в себя совокупность идей и комплекс математических средств, превратился в общенаучное средство исследования. Развитие понятия информации в современной науке привело к появлению её мировоззренческой, в особенности философской, интерпретаций. Однако, не существует одного общего определения понятия информации. Порождает дискуссию вопрос о предметной области понятия информации (является ли она свойством всех материальных объектов, или только живых и самоуправляющихся, или же только сознательных существ и т. п.). Проблема информации является одной из наиболее актуальных и фундаментальных в условиях современной научно-технической революции, характеризующейся, в частности, передачей информативных функций от человека к машинам в самых широких масштабах. Количественный подход – наиболее разработанная ветвь теории информации. Количество информации в двух сообщениях может быть совершенно одинаковым, а смысл совершенно разным. Два слова, например «Мир» и «Рим», содержат одинаковое количество информации, состоят из одних и тех же букв, но смысл слов различен. Оценка количества информации основывается на законах теории вероятностей, точнее, определяется через вероятности событий. Сообщение о событии, у которого только два одинаково возможных исхода, содержит одну единицу информации, называемую битом. Выбор единицы информации не случаен. Он связан с наиболее распространенным двоичным способом ее кодирования при передаче и обработке. Количество информации – это мера уменьшения неопределённости некоторой ситуации. Различные

количества информации передаются по каналам связи, и количество проходящей через канал информации не может быть больше его пропускной способности, т.е. количества информации проходящей здесь за единицу времени.

**1.3 Схема передачи информации.** Для повышения надежности в передаче и обработке информации вводят лишние символы - своеобразную защиту от искажений, которые не несут действительного содержания в сообщении (они избыточны). Упрощенная схема передачи информации представлена на рисунке 1. Рис. 1. Общая схема передачи информации.

**1.4 Кодирование информации.** Всякое событие, всякое явление служит источником информации и может быть выражено по-разному, разным способом, разным алфавитом. Чтобы информацию более точно и экономно передать по каналам связи, ее надо соответственно закодировать. Информация не может существовать без материального носителя, без передачи энергии. Закодированное сообщение приобретает вид сигналов-носителей информации. Они-то и идут по каналу. Выйдя на приемник, сигналы должны обрести вновь общепонятный вид. С этой целью сигналы пробегают декодирующее устройство, приобретая форму, удобную для абонента. Система связи сработала, цель достигнута. Когда говорят о каналах связи, о системах связи, чаще всего для примера берут телеграф. Но каналы связи - понятие очень широкое, включающее множество всяких систем, самых разных. Отдельные системы вычислительной машины передают одна другой информацию с помощью сигналов. Ведь вычислительная машина - автоматическое устройство для обработки информации, как станок - устройство для обработки металла. Машина не создает из «ничего» информацию, она преобразует только то, что в нее введено. Количественный метод - одно из направлений в теории информации - наиболее распространенный и наиболее разработанный. Существуют и другие методы. Они, в противоположность количественному, стараются ухватить смысл информации, ее ценность, ее качество, как, например, семантическая и прагматическая теории информации. Ценность информации можно определить как свойство информации, влияющей на поведение ее получателя. Физическими носителями информации являются **электронно-вычислительные машины, которые** появились в годы второй мировой войны. В 1943 году компания International Business Machines создала по заказу военно-морского флота США первый компьютер. Его сконструировали ученые Гарвардского университета и назвали "**Марк-1**". Его размеры 2 м в высоту и 15 м в длину. Созданный в Австралии в 1949 г. компьютер "**CSIR Mark-1**" выполнял самые разнообразные задачи для научного сообщества (от радиофизики до метеорологии) на протяжении долгих пятнадцати лет, вплоть до 1964 г. включительно. Появление компьютеров наряду с разнообразной информационной теорией создало мощную индустриальную основу для возникновения общества, в котором информация как предмет труда играет все более заметную роль. По некоторым подсчетам, суммарный информационный поток возрастает в среднем пропорционально квадрату промышленного потенциала. Иначе говоря, увеличение вдвое производительных сил страны потребует 4-кратного увеличения потока информации. Все очевиднее возрастающая зависимость национального развития от состояния информационных ресурсов. Чтобы быть в курсе новейших научно - технических веяний, как признают сами американцы, сегодня "нужно иметь информацию едва ли не по всем важнейшим исследованиям, которые ведутся за рубежом". Информационное невежество в наши дни ведет к технологическому банкротству и национальному унижению, к превращению страны в сырьевую колонию или свалку устаревшей технологии. С середины 70-ых годов возникло понятие "информационное общество". Требованием этого общества, его лозунгом стало "Как можно больше информации и как можно быстрее!" Ускоряющееся течение жизни, интенсификация процесса производства в результате автоматизации, быстрое накопление знаний - все это повышает экономическую значимость информации.

**1.5 Основные определения темы. 1. Архитектура информационной системы** - концепция, определяющая модель, структуру,

выполняемые функции и взаимосвязь компонентов информационной системы.

2. **Данные** - сведения: полученные путем измерения, наблюдения, логических или арифметических операций; и представленные в форме, пригодной для постоянного хранения, передачи и (автоматизированной) обработки.

3. **Документная информация** - информация, содержащаяся в документах.

4. **Защищенность информационной системы (Security)** - способность системы противостоять несанкционированному доступу к конфиденциальной информации, ее искажению или разрушению. Различают два аспекта защищенности: -1- техническую защиту (свойство недоступности); и -2- социальную защищенность (свойство конфиденциальности).

5. **Знак** - материально, чувственно воспринимаемый предмет, явление или действие, служащие: - для обозначения другого предмета, свойства или отношения; - для переработки и передачи информации. Определяющей стороной знака является его значение, которое может быть предметным, смысловым или экспрессивным. Различают языковые и неязыковые знаки.

6. **Измерительная информация (Measurement information)** - информация о значениях физических величин. В измерительную информацию могут также входить данные о погрешности измерений, о числе измерений и др.

7. **Информатика (Informatics)** - отрасль знаний, изучающая общие свойства и структуру научной информации, а также закономерности и принципы ее создания, преобразования, накопления, передачи и использования в различных областях человеческой деятельности.

8. **Информационная система** - по законодательству РФ - организационно упорядоченная совокупность документов (массивов документов) и информационных технологий, в том числе с использованием средств вычислительной техники и связи, реализующих информационные процессы. Информационные системы предназначена для хранения, обработки, поиска, распространения, передачи и предоставления информации.

9. **Информационная услуга** - услуга, ориентированные на удовлетворение информационных потребностей пользователей путем предоставления информационных продуктов. Информационные услуги - по законодательству РФ - действия субъектов (собственников и владельцев) по обеспечению пользователей информационными продуктами.

10. **Информационное наполнение (Content)** - тексты, графика, мультимедиа и иное информационно значимое наполнение информационной системы. Существенными параметрами информационного наполнения являются его объем, актуальность и релевантность. В сети Internet информационное наполнение организуется в виде веб-страниц средствами языка гипертекстовой разметки.

11. **Информационное общество (Information society)** - концепция постиндустриального общества; новая историческая фаза развития цивилизации, в которой главными продуктами производства являются информация и знания. Отличительными чертами информационного общества являются: - увеличение роли информации и знаний в жизни общества; - возрастание доли информационных коммуникаций, продуктов и услуг в валовом внутреннем продукте; - создание глобального информационного пространства, обеспечивающего (а) эффективное информационное взаимодействие людей, (б) их доступ к мировым информационным ресурсам и (в) удовлетворение их потребностей в информационных продуктах и услугах.

12. **Информационное пространство** - совокупность (1) банков и баз данных, (2) технологий их сопровождения и использования, (3) информационных телекоммуникационных систем, функционирующих на основе общих принципов и обеспечивающих: - информационное взаимодействие организаций и граждан; и - удовлетворение их информационных потребностей. Основными компонентами информационного пространства являются: (1) информационные ресурсы, (2) средства информационного взаимодействия и (3) информационная инфраструктура.

13. **Информационные процессы** - по законодательству РФ - процессы сбора, обработки, накопления, хранения, поиска и распространения информации.

14. **Информационный массив** - совокупность зафиксированной информации, предназначенная для хранения и использования и рассматриваемая как единое целое.

Информация может быть зафиксирована в виде публикаций, отчетов, электронных записей, микрокопий и т.д. Обычно на предприятиях и в учреждениях информационные массивы формируются по функциональному признаку.

15. **Информационный подход** (Information approach) - метод научного познания объектов, процессов или явлений природы и общества, согласно которому в первую очередь выявляются и анализируются наиболее характерные информационные аспекты, определяющие функционирование и развитие изучаемых объектов.

16. **Информация** - по законодательству РФ - сведения о лицах, предметах, фактах, событиях, явлениях и процессах независимо от формы их представления. Информация уменьшает степень неопределенности, неполноту знаний о лицах, предметах, событиях и т.д. Её свойства:

1. **качество** (Information quality) - совокупность свойств, отражающих степень пригодности конкретной информации об объектах и их взаимосвязях для достижения целей, стоящих перед пользователем;
2. **адекватность информации** - уровень соответствия создаваемого с помощью полученной информации образа реальному объекту, процессу, явлению и т.п.;
3. **достоверность информации** (Validity of information) - свойство информации быть правильно воспринятой. В общем случае достоверность информации достигается: - указанием времени свершения событий, сведения о которых передаются;- сопоставлением данных, полученных из различных источников;- своевременным вскрытием дезинформации; - исключением искаженной информации и др.
4. **ценность информации** - свойство информации, изменяющее поведение её получателя.
5. **безопасность информации** (Information security) - состояние защищенности информационной среды общества. Информационная безопасность имеет три основные составляющие: -1- конфиденциальность - защита чувствительной информации от несанкционированного доступа; -2- целостность - защита точности и полноты информации и программного обеспечения; -3- доступность - обеспечение доступности информации и основных услуг для пользователя в нужное для него время.

17. **Пользователь информационной системы** (Information system user) - лицо, группа лиц или организация, пользующееся услугами информационной системы для получения информации или решения других задач.

18. **Экономика, основанная на знаниях** (Information economy; Knowledge economy) - экономика, в которой большая часть валового внутреннего продукта обеспечивается деятельностью по производству, обработке, хранению и распространению информации и знаний, причем в этой деятельности участвуют более половины занятых.

19. **Энтропия** (Entropy) - мера внутренней неупорядоченности информационной системы. Энтропия увеличивается при хаотическом распределении информационных ресурсов и уменьшается при их упорядочении.

20. **Эффективное кодирование информации** - уменьшение длины кодового слова для часто встречающихся символов и увеличение для редко встречающихся.

**2. ДОКУМЕНТАЛЬНЫЕ ИНФОРМАЦИОННО-ПОИСКОВЫЕ СИСТЕМЫ (ДИПС).**

**2.1 Теория ДИПС.** Документальная информационно-поисковая система (ДИПС) - информационно-поисковая система, предназначенная для отыскания документов, содержащих необходимую информацию. Поисковый массив ДИПС состоит из поисковых образов документов или из самих документов. В отличие от традиционных баз данных, ориентированных на полное и точное представление данных достаточно простой смысловой структуры, документальные базы данных ориентированы на частичное, приближенное представление данных, имеющих значительно более сложную смысловую структуру, представленных на входе в форме текста. Основная функция ДИПС – информационное обеспечение потребителей на основе выдачи ответов на их запросы. Осуществление выдачи системой требуемых данных реализуется с помощью главной операции ДИПС – проведения информационного поиска. Информационный поиск – процедура отыскания документа, содержащего ответ на

заданные потребителем вопросы. Информационный поиск в системе проводится на основе поступившего от потребителя информационного запроса в результате информационной потребности. Информационный запрос является частным проявлением информационной потребности. Информационная потребность – это потребность человека в определённой информации в процессе его практической деятельности. Информационный запрос (information demand, query) – частное значение информационной потребности потребителя в определённые моменты времени, выраженные на естественном языке (ЕЯ). Однако, запрос может быть неправильно сформулирован потребителем и не отражать его истинной информационной потребности в момент обращения к системе. Таким образом, при проведении информационного поиска в системе, фактически рассматривается не информационная потребность пользователя, а только информационный запрос, в ответ на который и выдаются те или иные документы системы. Следовательно, реакцию системы необходимо рассматривать не только по отношению к информационной потребности, но и по отношению к информационному запросу. Для выражения данных отношений в теории ДИПС введены два фундаментальных понятия: пертинентность и релевантность. **Пертинентность** – соответствие смыслового содержания документа информационной потребности потребителя. Документы, содержание которых удовлетворяет информационной потребности, называют пертинентными. **Релевантность** – соответствие содержания документа информационному запросу в том виде, в каком он сформулирован, а документы, содержание которых отвечает запросу потребителя, носят название релевантных. Автоматизация процесса информационного поиска потребовала формализации основного смыслового содержания информационного запроса и документов в виде соответственно поискового предписания (ПП) и поисковых образов документов (ПОД). Для записи ПП и ПОД применяются специальные языки, называемые информационно-поисковыми (или просто информационными). В процессе проведения информационного поиска в ДИПС определяется степень соответствия содержания документов и запроса пользователя путём сопоставления ПОД с ПП. А на основе такого сопоставления принимается решение о выдаче документа (он признаётся релевантным) или его невыдаче (он считается нерелевантным). Решение делается на основании критерия смыслового соответствия (КСС). Критерий смыслового соответствия – это набор некоторых правил ДИПС по которым определяется смысловая близость между ПОД и ПП. Критерий может задаваться 1. явно и 2. неявно. КСС базируется на фактической релевантности. Фактическая релевантность – это соответствие содержания документа информационному запросу и может быть установлена только человеком в процессе осмысления содержания документа и запроса. Близкий к КСС и информационному запросу термин ключевое слово. Ключевые слова в контексте (Keyword-in-context) - методология автоматизированного поиска, используемая для создания указателей текстов или списка заголовков документов, при которой каждое ключевое слово хранится вместе с окружающим его текстом.

**2.2 Общая функциональная структура ДИПС.** В состав типичной ДИПС входят четыре подсистемы: 1. Подсистема ввода и регистрации 2. Подсистема обработки 3. Подсистема хранения 4. Подсистема поиска. Подсистема ввода и регистрации решает следующие основные задачи: - создание электронных копий бумажных документов; например, сканирование с последующим распознаванием текста или ввод с клавиатуры; - обеспечение подключения к каналам доставки электронных документов; - распознавание, а при необходимости и преобразование формата электронных документов; - присвоение электронным документам уникальных идентификаторов (регистрация), а также введение таблицы синхронизации имён (при необходимости сохранения прежних имён). Далее все документы поступают в подсистему хранения, где информация хранится в сжатом виде. Далее документ поступает в подсистему обработки. Её функция - формирование

ПОД, который сохраняется в индексе. **2.3 Представление смыслового содержания текста.** Недостатки естественного языка (ЕЯ). 1. Многообразие средств передачи смысла- лексика- контекст- парадигматические отношения между словами- текстуальные отношения между словами- ссылки на слова (словосочетания, фразы и т.д.) ранее упоминавшиеся в тексте сообщения. 2. Семантическая неоднозначность. 3. Синонимия - тождественность. 4. Многозначность: - Полисемия – совпадение названий различных предметов, имеющих между собой какие-либо общие свойства и признаки; например, слово «команда»; - омонимия – совпадение названий различных предметов, не имеющих общего смысла - «ЛУК», «КЛЮЧ». 5. Эллипсность – пропуски подразумеваемых слов. **2.4 Информационно-поисковые языки (ИПЯ).** Невозможность использования ЕЯ в качестве основного средства представления информации в ДИПС приводит к необходимости применения искусственных языковых средств. ИПЯ – специализированный искусственный язык, предназначенный для описания основного смыслового содержания поступающих в систему сообщений, с целью обеспечения возможности последующего их поиска. Различают ИПЯ: 1. классификационные (предкоординируемые) языки 2. дескрипторные (посткоординируемые) языки - с грамматикой - без грамматики - с контролируемой лексикой - со свободной лексикой. Разница между ними в процедуре построения предложений из лексических единиц. В первом случае предложение строится так: Политика. Внутренняя. Федеральная..... Региональная и т.д. и т.п. Частный случай это рубрикатор. Рубрикатор – ориентированный граф, состоящий из независимых деревьев. «Листья» деревьев называются рубриками – объектами, инкапсулирующими знания о конкретных фрагментах данной предметной области. Рубрикатор формируется экспертами в данной области знаний. В дескрипторных языках лексические единицы никак заранее не связаны. **2.5 Обработка входящей текстовой информации.** Различают рубрицирование и индексирование. Рубрицирование - перевод документов, поступающих в ДИПС с ЕЯ на ИПЯ с помощью применения дескрипторного языка. Индексирование - перевод документов, поступающих в ДИПС с ЕЯ на ИПЯ с помощью применения классификационного языка. В процессе перевода необходима нормализация - приведение к канонической форме. Различают нормализацию слов и словосочетаний. Нормализованные слова и словосочетания часто называют терминами. Лингвистический анализ состоит из двух этапов: 1. морфологический анализ 2. синтаксический анализ. Суть морфологического анализа в сопоставлении слов со словарём. Далее следует синтаксический анализ. Его суть в сравнении предложений со словарём. Далее следует или автоматическое индексирование - присваивание термину индекса на основе частоты встречаемости его в том или ином тексте или, автоматическое рубрицирование – разбиение текста по заранее выделенным рубрикам. Различают автоматическое индексирование: 1. Основанное на знаниях 2. Основанное на обучении на примерах. **2.6 Основные определения темы.** 1. **Документальная информационно-поисковая система** (Document information retrieval system) - информационно-поисковая система, предназначенная для отыскания документов, содержащих необходимую информацию. Поисковый массив документальной ИПС состоит из поисковых образов документов или из самих документов. 2. **Информационно-поисковая система** (Information retrieval system) - система, выполняющая функции: - хранения больших объемов информации; - быстрого поиска требуемой информации; - добавления, удаления и изменения хранимой информации; - вывода информации в удобном для человека виде. Различают: - автоматизированные (computerised); - библиографические (reference); - диалоговые (online); - документальные и фактографические информационно-поисковые системы (ДИПС и ФИПС). 3. **Фактографическая информационно-поисковая система** (Factographic information retrieval system) - **информационно-поисковая система, обеспечивающая выдачу непосредственно фактических сведений, затребованных потребителем в информационном запросе. Поисковый массив**

**фактографической ИПС состоит из описаний фактов, извлеченных из документов и представленных на некотором формальном языке. 3. ДИПС. 3.1 Автоматическое индексирование.**

Автоматическое индексирование (АИ) документов может основываться на простых, однословных или многословных составных терминах (фразах). Простые, однословные термины далеко не идеальны для индексирования, поскольку смысл слов вне контекста нередко бывает неоднозначным. Термины-фразы более осмысленны, обладают большей дискриминирующей мощностью. Для генерации фраз может использоваться как синтаксический анализ, так и ряд эвристических алгоритмов. Ниже мы рассмотрим один из них. Предположим, что термин-фраза состоит из основы фразы (обычно это её главная часть) и остальных компонентов. Термин с частотой вхождения в документы, превышающей установленный порог, например  $df > 2$ , отмечается как основа фразы. Другими компонентами фразы должны быть термины со средней или низкой частотой вхождения. При этом учитывается их связь с основой фразы, например, размещение их в одном предложении или на некотором заданном расстоянии друг от друга. Для генерации групп взаимосвязанных слов по замеченным закономерностям совместного их вхождения в документы применяются методы группирования или кластеризации терминов. Если представить матрицу терминов-документов в виде двухмерного массива, то вышеупомянутый метод сравнивает друг с другом столбцы матрицы и делает заключение о том, входит та или иная группа терминов в несколько документов совокупности. Если такое неоднократное вхождение имеет место, то термины считаются связанными и группируются в один класс. Простые и составные термины, выполняющие чисто грамматическую функцию, заносятся в так называемые списки исключения и удаляются. Основу современных методов автоматического индексирования составляет присваивание весовых коэффициентов терминам на основе статистических характеристик. Предположим, что в исследуемой совокупности имеется  $N$  документов. Пусть  $tf_{ij}$  – частота вхождения термина  $T_j$  в документ  $D_i$ . Индексирование на основе частоты термина позволяет достичь лишь одной из целей индексирования – полноты поиска. Между тем термины, сконцентрированные в отдельных документах совокупности, можно использовать для повышения точности поиска. Это позволит отделить документы, где такие термины встречаются, от тех, где их нет. Пусть  $df$  – число документов, в которых встречается термин  $T_j$ . Тогда величина  $\log(N/df_j)$  может служить хорошим индикатором того, является ли термин  $T_j$  дискриминатором документов. Частоту термина и полученную выше величину можно объединить в рамках единой модели индексирования по частоте (здесь  $w_{ij}$  обозначает вес термина  $T_j$  в документе  $D_i$ ).  $w_{ij} = tf_{ij} \log(N/df_j)$  Ещё один статистический метод индексирования основывается на дискриминации по термину. Здесь каждый документ рассматривается как точка в пространстве документов. Чем больше сходства у множеств терминов двух документов, тем ближе расположены соответствующие точки в пространстве документов (иными словами, повышается плотность точек в пространстве документов), и наоборот. В рамках данной схемы можно оценивать качество термина как дискриминатора документа, основываясь на том, какие изменения произойдут в пространстве документов после введения термина в индекс. Для количественной оценки такого изменения удобно использовать увеличение или уменьшение расстояния между документами. Термин является хорошим дискриминатором, если его введение увеличивает среднее расстояние между документами. Другими словами, термин с хорошими дискриминирующими качествами снижает плотность в пространстве документов. Дискриминирующая характеристика термина  $T_j$  обозначаемая  $dv_j$ , вычисляется как разность между плотностями пространства документов до и после введения термина  $T_j$ . Оказалось, что часто встречающиеся термины имеют отрицательные значения дискриминирующих характеристик, термины со средней частотой – положительные, а для редко встречающихся терминов эти значения близки к нулю. Для совместного учёта

частоты термина и его дискриминирующей характеристики применяют схему взвешивания, основанную на выражении  $w_{ij} = tf_{ij} \cdot dv_j$ . Полученные значения весов терминов могут использоваться в процессе принятия решения о включении каждого из терминов в ПОД. Однако чаще решение не принимается, а в ПОД заносятся все термины, встретившиеся в документе, и их веса.

**3.2 Автоматическое рубрицирование.** Различают автоматическое рубрицирование (АР): 1. Основанное на знаниях; 2. Основанное на обучении по примерам. В системах, реализующих методы автоматического рубрицирования, основанные на знаниях, используются заранее сформированные базы знаний, в которых описываются языковые выражения, соответствующие той или иной рубрике, правила выбора между рубриками. Процесс создания подобных систем часто сравнивают с созданием экспертных систем для диагностики и классификации. Наибольшее распространение среди данных методов получили две модели представления знаний: 1. модель семантической сети 2. продукционная модель. В первом случае знания о предметной области описываются независимо от рубрикатора в специальном виде тезауруса, который связывается с одним или более рубрикаторами гибкой системой отношений. Существует несколько определений термина «тезаурус». **Под тезаурусом** (от греч. thesauros – запас, сокровище) понимается иерархическая сеть понятий и отношений между ними. Тезаурус может быть разработан независимо от какой-либо системы рубрицирования. В нём могут быть накоплены разнообразные варианты представления в тексте понятий предметной области (дескрипторов). В качестве вариантов (синонимов или эквивалентов) дескрипторов в тезаурусе встречаются именные и глагольные группы, отдельные существительные, прилагательные или глаголы. Информационно-поисковый тезаурус (ИПТ) – словарь дескрипторного информационно-поискового языка с зафиксированными в нём парадигматическими отношениями лексических единиц. Тезаурус содержит список ключевых слов, которыми может быть охарактеризовано содержание документов, с выделением слов, рекомендованных для индексирования (дескрипторов). Парадигматические отношения указывают общность или противопоставление значений и использования лексических единиц. Обычно информационно-поисковые тезаурусы оформляются в виде книг. В соответствии с тематическим профилем различают многоотраслевые, отраслевые и узкотематические тезаурусы. Тезаурус – термин, означающий "весь запас сведений некоторой системы, сгруппированных по заданным признакам и с выявлением связей между элементами системы". Тезаурус может быть разработан в полуавтоматическом режиме. Например, сначала обрабатываются совокупность документов большого объёма при помощи программ морфологического и синтаксического анализа с целью выделения терминоподобных групп слов. Затем выбранные группы слов исследуются экспертами и принимается решение относительно того: – Может ли данная группа слов включена в тезаурус (в этом случае она становится термином); – Является ли данный термин дескриптором или синонимом другого дескриптора; – Как должны быть описаны отношения данного термина. Кроме того, в комплекс знаний могут также быть включены дополнительные базы данных, например: географическая база данных, содержащая описание географических объектов, База данных организаций, персоналий и т.д. Тезаурус и базы данных имеют одну структуру и состоят из следующих частей: 1. Дескрипторов, которые соответствуют понятиям или конкретным объектам. Обычно дескриптором является существительное или именная группа. 2. Каждый дескриптор имеет совокупность текстовых входов или синонимов. Текстовый вход может быть существительным, прилагательным или группой существительного. Одно слово может быть синонимом различных дескрипторов. Устранение смысловой неоднозначности производится во время автоматической обработки документа. 3. Отношения между дескрипторами внутри каждой базы данных, например: – Более широкий термин (выше) – Более узкий термин (ниже) – Связанный термин (ассоциация) – Целое для термина (часть) – Часть для термина (целое) 4. Отношения между дескрипторами различных баз данных. В данном случае добавляется

отношение – «равенство термина», которое появляется, когда базы данных содержат дескрипторы, соответствующие одному понятию или объекту. Дескриптор D1 находится в дескрипторной среде дескриптора D, если между D1 и D существует дескрипторное отношение или существует транзитивная зависимость. Дескриптор D называют главным дескриптором среды. Транзитивная зависимость – это такая зависимость, при которой неключевой атрибут зависит от одного или более неключевых атрибутов. Иерархическая организованность тезауруса и наличие тезаурусных связей позволяет использовать понятия среды дескрипторов и главных дескрипторов (опорных дескрипторов) среды для формирования дескрипторных кустов, используемых при автоматическом рубрицировании текстов в данной технологии. В целом же комплекс знаний представляет собой иерархическую сеть, полнота и целостность которой поддерживается и отслеживается экспертами. Существует два типа представления рубрик последовательностью опорных дескрипторов в виде булевских нормальных форм: Дизъюнкция опорных дескрипторов  $D1 \vee D2 \vee \dots \vee Dn$  – Конъюнкция дизъюнкций опорных дескрипторов  $(D11 \vee D12 \vee \dots \vee D1n) \& \dots \& (Dm1 \vee Dm2 \vee \dots \vee Dmk)$ . Для каждой рубрики рубрикатора может быть выбран свой тип представления. После того, как для всех рубрик рубрикатора установлены связи с соответствующими опорными дескрипторами, автоматически определяются рубрики для всех дескрипторов тезауруса. Таким образом, для каждого дескриптора создаётся список соответствующих рубрик с указанием того, в какую из дизъюнкций рубрики входит данный дескриптор. Каждая рубрика в данной технологии фиксирует запрос пользователя, который описывается посредством дескрипторов тезауруса. Дальнейшее развитие технологии состоит в предоставлении пользователю возможности описывать рубрику на ЕЯ. Суть процесса рубрицирования в рамках данного подхода состоит в выделении из текста опорных дескрипторов и отношений между ними с последующим сопоставлением их с описаниями рубрик. Однако эта технология очень трудоёмкая и требует больших временных затрат (от нескольких месяцев до нескольких лет). Итак, это была описана модель семантической сети. Теперь перейдём к описанию метода продукционной модели, основу которой составляет выделение из текста концепций (или понятий), заранее описанных экспертом. Каждое понятие предметной области описывается экспертом при помощи особой конструкции – определения понятия, объединяющего в себе набор характерных для данного понятия слов и фраз. Определение понятия – выражение, записанное на специальном языке, позволяющем объединять эти слова и фразы при помощи стандартных булевых функций. В определении понятия при записи слов и фраз допускается использование символов-шаблонов (&, \* и т.д.), что позволяет отказаться от процедуры морфологического анализа, используемой для нормализации лексики документа. Например, определение понятия золото:  $(gold(\&n(reserve!medal!jewelry)))$  Процесс рубрицирования разбивается на два этапа. Первый – выделение понятий из текста. Второй – принятие решения о принадлежности текста к конкретной рубрике. Решение принимается на основе правил рубрицирования. Правила рубрицирования формулируются экспертом заранее. Совокупность определений понятий и правил рубрицирования составляет базу правил. Разработка базы правил – очень трудоёмкий процесс, требующий привлечения высококвалифицированных специалистов как в предметной области, так и в области инженерии знаний. Преимущества данного подхода: 1. высокое качество рубрицирования 2. высокое быстродействие на тех потоках, для которых они проектировались. Недостатки: 1. Высокая трудоёмкость и значительные затраты 2. узость применения. Системы АР, основанные на этом методе обучения по примерам рассматривают в качестве понятий, которым нужно обучиться, рубрики. Выделяют методы: 1. статистического рубрицирования; 2. нейросетевого рубрицирования. Статистическое рубрицирование – определение степени соответствия терминологического портрета документа и терминологического портрета рубрик на основе статистических характеристик субъектов сравнения. Терминологический портрет документа – совокупность

наиболее важных терминов, содержащихся в тексте документа. Важность термина – это частота его встречаемости. Терминологический портрет рубрики – набор наиболее характерных для этой рубрики терминов с их весами. Формирование терминологических портретов каждой рубрики производится экспертом не вручную, а с помощью одной из технологий обучения рубрикатора. При этом роль эксперта сводится к формированию для каждой рубрики обучающей выборки. Обучающая выборка – совокупность максимально коротких фрагментов текстов, содержащих полное и минимально избыточное лингвистическое наполнение одной обучаемой рубрики. Выделение терминологических терминов для рубрики производится автоматически, на основе их весов, которые могут быть получены в процессе анализа обучающей выборки. Например:  $w_{it} = \log(N_t/df_{it})$  где  $N_t$  – количество документов в обучающей выборке, принадлежащих рубрике,  $df_{it}$  – количество документов в обучающей выборке, содержащих термин. Список характеристических терминов рубрики упорядочен по убыванию весов терминов в ней. Достоинства: 1. простота определения семантики рубрики, что даёт возможность организовать автоматическое обучение рубрик 2. универсальность подхода, заключающаяся в том, что таким способом может быть определена семантика очень широкого класса рубрик из любой предметной области 3. наличие аппарата количественной оценки релевантности документа рубрикам 4. высокое быстродействие. Недостаток – очень низкое качество рубрицирования. Нейросетевой метод – использует нейронную сеть (НС) в качестве обучаемого классификатора. Достоинства НС: 1. Строгое математическое обоснование 2. огромное, (в тысячу раз большее) по сравнению с другими нейросетевыми парадигмами, быстродействие. Недостатки НС: 1. Экспертам непонятно, как НС работает 2. на обучение сети требуется очень много времени. **3.3**

**Поиск текстовой информации.** Модель поиска характеризуется 4 параметрами 1. представлением документов и запросов 2. критерием смыслового содержания 3. методами ранжирования результатов запроса 4. механизмами обратной связи, обеспечивающей оценку релевантности пользователем. Наиболее распространённая модель, отвечающая, как минимум первым трём параметрам, это булева модель. Она наиболее лёгкая. При наличии термина в документе соответствующая переменная принимает значение True. Мерой соответствия запроса документу служит значение статуса выборки (RSV, retrieval status value). Применяется во многих коммерческих системах. Вторая модель – это модель нечётких множеств. Основывается на теории частичной принадлежности элемента тому или иному множеству. Третья модель. Пространственно-векторная модель – основана на предположении, что совокупность документов можно представить набором векторов в пространстве, определяемом базисом, из  $n$  нормализованных векторов терминов. Четвёртая модель. Вероятностные модели. **3.4 Методы ведения обратной связи с пользователем.**

В отличие от среды баз данных в ДИПС нет чёткого представления документов и пользовательских запросов. Пользователи обычно начинают с неточного и неполного запроса, а, следовательно – с низкой эффективности поиска, постепенно уточняя его методом итераций. Множество документов, считающихся релевантными, формируют положительную обратную связь, а множество документов, рассматриваемых как нерелевантные, – отрицательную. Существуют два основных подхода к использованию такой обратной связи: 1. Модификация запроса 2. Модификация представления документов. **3.5 Оценка качества ДИПС.** Любой ДИПС присущи два основных типа ошибок: - ошибки первого рода (или пропуск цели): невыдача потребителю фактически релевантных его запросу документов - ошибки второго рода (или ложная тревога, иначе шум): выдача потребителю нерелевантных документов, которые не отвечают поставленному запросу. Наличие ошибок в реальной системе обуславливает разбиение всего массива документов на 4 подмассива (Таблица 1). Таблица 1. Разбиение массива документов

| Выданные                                | Невыданные | Релевантные                              | АС | Нерелевантные | BDA |
|-----------------------------------------|------------|------------------------------------------|----|---------------|-----|
| массив выданных релевантных документов; | B          | массив выданных нерелевантных документов | C  |               |     |

- массив невыданных релевантных документов  $D$  - массив невыданных нерелевантных документов. Введём следующие обозначения:  $a$  – количество выданных релевантных документов;  $b$  - количество выданных нерелевантных документов;  $c$  - количество невыданных релевантных документов;  $d$  - количество невыданных нерелевантных документов. Существуют следующие показатели эффективности ДИПС: 1. Коэффициент полноты  $p$ , характеризующий долю выданных релевантных документов во всём массиве релевантных документов:  $p = a/a + c$  2. Коэффициент точности  $n$ , характеризующий долю выданных релевантных документов во всём массиве выданных документов:  $N = a/a + b$  3. Коэффициент шума  $e$ , характеризующий долю выданных нерелевантных документов во всём массиве выданных документов:  $E = b/a + b = 1 - n$  4. Коэффициент осадка  $q$ , характеризующий долю выданных нерелевантных документов во всём массиве нерелевантных документов:  $Q = b/b + d$  5. Коэффициент специфичности  $k$ , характеризующий долю невыданных нерелевантных документов во всём массиве нерелевантных документов:  $K = d/b + d$  Для удобства перечисленные показатели измеряют в процентах. Также для оценки используют: Быстродействие ДИПС (интервал времени между моментом формулировки запроса и получением ответа на него)- Пропускная способность (количество вводимых документов и количество ответов в единицу времени при заданных значениях коэффициентов полноты и точности)- Производительность (оценивается количеством пользователей системы и частотой обращения с их стороны)- Надёжность оценивается вероятностью того, что система будет выполнять свои функции при заданных условиях в течение требуемого времени)- Тип запросов, обслуживаемых системой.

**3.6 Основные определения темы.** 1. Ассоциативный дескриптор (**Related term**) - дескриптор, связанный с другим дескриптором семантической связью, характер которой не указывается. 2. Видовой дескриптор (**Specific term**) - нижестоящий дескриптор, обозначающий видовое понятие. 3. Вспомогательный дескриптор (**Квалификатор Auxiliary descriptor**) - дескриптор, который используют только совместно с другими дескрипторами. 4. Вышестоящий дескриптор (**Широкий дескриптор Broader term**) - дескриптор, обозначающий либо родовое понятие, либо целое по отношению к данному дескриптору, обозначающему часть этого целого. 5. Графический указатель информационно-поискового тезауруса (**Graphic display of a thesaurus**) - совокупность семантических схем, в графической форме представляющих парадигматические отношения между дескрипторами с помощью диаграмм, стрелок и т.п. 6. Дескриптор (**Descriptor**) - лексическая единица: - выраженная информативным словом (вербально) или кодом; и - являющаяся именем класса синонимичных или близких по смыслу ключевых слов. Многозначному слову естественного языка соответствует несколько дескрипторов, а несколькими синонимичным словам и выражениям - один дескриптор. Дескрипторы используются для координатного индексирования документов и информационных запросов с целью последующего поиска. 7. Дескрипторный информационно-поисковый язык (**Информационно-поисковый язык дескрипторного типа; Дескрипторный язык Descriptor language**) - информационно-поисковый язык, предназначенный для координатного индексирования документов и информационных запросов посредством дескрипторов и/или ключевых слов. 8. Дескрипторный словарь - словарь дескрипторного информационно-поискового языка, в котором приведены в общем алфавитном ряду дескрипторы и их синонимы без указания других отношений лексических единиц. Дескрипторный словарь является упрощённым вариантом информационно-поискового тезауруса, в котором зафиксированы (преимущественно) синонимические связи. 9. Идентификатор (**Identifier**) - в ИПС - имя собственное, используемое как дескриптор. 10. Иерархический указатель информационно-поискового тезауруса (**Hierarchical display of a thesaurus**) - список

дескрипторов высшего уровня иерархии, в котором для каждого из дескрипторов приводятся подчиненные нижестоящие дескрипторы, расположенные в порядке убывания общности. 11. Код дескриптора (Concept symbol) - код, используемый в информационно-поисковом тезаурусе для представления эквивалентных дескрипторов и их синонимов. 12. Лексикографическая база данных (Lexicographic database) - база данных, запись в которой содержит данные об одной лексической единице и соответствует статье словаря. 13. Лексико-семантический указатель - основная часть информационно-поискового тезауруса, в которой в едином алфавитном порядке перечислены все дескрипторы и нондескрипторы с указанием их парадигматических отношений. 14. Лексическая единица - слово, устойчивое словосочетание или другая единица языка, способная обозначать предметы, явления, их признаки и т.п. 15. Нижестоящий дескриптор (Узкий дескриптор Narrower term) - дескриптор, обозначающий либо видовое понятие, либо часть по отношению к понятию, представленному вышестоящим дескриптором. 16. Нондескриптор (Недескриптор; Аскриптор Non-descriptor) - лексическая единица в информационно-поисковом тезаурусе, которая не может быть использована для координатного индексирования и подлежит замене одним или несколькими заменяющими ее дескрипторами. 17. Объединяющий дескриптор (Collective term) - вышестоящий дескриптор, обозначающий целое в отношении часть-целое. 18. Определение (Definition) - установление смысла незнакомого термина (слова) при помощи знакомых и уже осмысленных терминов. 19. Партитивный дескриптор (Partitive term) - нижестоящий дескриптор, представляющий часть или элемент в отношении часть-целое. 20. Предпочтительный термин (Preferred term) - термин, выбранный для использования в качестве дескриптора из множества эквивалентных по смыслу терминов. 21. Родовой дескриптор (Generic term) Родовой дескриптор - вышестоящий дескриптор, обозначающий родовое понятие. 22. Семантика (Semantics От греч. Semantikos – обозначающий) Семантика - раздел языкознания и логики, исследующий проблемы, связанные со смыслом, значением и интерпретацией лексических единиц. 23. Семантический информационный барьер (Терминологический информационный барьер) - информационный барьер, обусловленный несовпадением толкований одних и тех же слов, терминов и символов разными людьми. 24. Тезаурусная статья (Thesaurus entry) - элемент лексико-семантического указателя, состоящий из дескриптора или нондескриптора и всей, относящейся к нему информации. 25. Термин (Term от лат. Terminus – граница) - слово или словосочетание, являющееся точным обозначением определенного понятия какой-либо области знания. В тексте, предназначенном для перевода, термин выступает как единица перевода.

**4. ХАРАКТЕРИСТИКА МИРОВЫХ ИНФОРМАЦИОННЫХ РЕСУРСОВ. 4.1 Сети и протоколы, информационное обслуживание.** Компьютерная сеть - это объединение автономных персональных компьютеров для совместного использования вычислительных ресурсов (процессора, памяти и периферии - например, дорогостоящего лазерного принтера). Компьютерную сеть в пределах сравнительно небольшой территории называют локальной. Сеть, охватывающую большие пространства, - глобальной. Локальная сеть обычно организуется и работает в пределах одной фирмы (организации) и объединяет компьютеры на рабочих местах для более быстрого и качественного обмена информацией. Каждая организация, эксплуатирующая более десятка персональных компьютеров (ПК), старается объединить их в локальную сеть с целью уменьшения бумажного документооборота и повышения эффективности деятельности своих подразделений. Естественно, что почти каждая такая сеть должна иметь выход на внешних заказчиков. Глобальная сеть — это чаще всего самостоятельная (в техническом и юридическом отношении) структура

и другие фирмы подключаются к ней для работы за определённую плату. Одна из первых компьютерных сетей Arpanet была создана почти 30 лет тому назад по заказу исследовательского подразделения министерства обороны США. Поначалу эта сеть объединяла несколько мощных ЭВМ в организациях военно-промышленного комплекса, учебных и научно-исследовательских центрах. Arpanet располагала достаточно большим количеством терминалов пользователи которых могли выполнять свои программы на удалённых компьютерах. Оперативный характер такой деятельности и возможность доступа к высокопроизводительным процессорам предопределили успех вычислительных сетей. В настоящее время число крупных (глобальных) компьютерных сетей перевалило за 50000, а количество небольших (локальных) вычислительных сетей исчисляется сотнями тысяч. На первых порах, когда разные сети базировались на разных вычислительных платформах, т.е. использовали несовместимые технические и программные средства, когда еще не были приняты стандарты на передачу данных, междосетевое общение было затруднено. Серьёзный сдвиг в преодолении этих барьеров произошел в 1974 году, когда был разработан междосетевой протокол обмена информацией TCP/IP (Transmission Control Protocol/Internet Protocol). Однако прошло ещё почти 10 лет, пока этот протокол не был принят в обязательном порядке на всех узловых компьютерах сети Arpanet. К тому моменту к Arpanet присоединилось несколько сотен других сетей, и набор услуг вычислительного характера начал постепенно вытесняться услугами информационного плана. Протокол TCP/IP не является единственным стандартом на передачу данных в сетях. Поэтому между двумя сетями, использующими для внутреннего общения разные протоколы, организуются серверы-посредники, называемые шлюзами. Любой протокол обмена информацией между двумя абонентами представляет собой набор соглашений, в которых оговариваются различные технические характеристики и программные средства, участвующие в приёме / передаче данных. В состав этих характеристик входят скорость передачи данных, средства контроля и устранения ошибок, форматы заголовков и способы адресации абонентов, методы сжатия и распаковки информации и многое другое. Часть таких соглашений обеспечивается аппаратными возможностями модемов, другие - обслуживающими программами. Два ПК, оборудованные модемами и снабжённые соответствующим программным обеспечением, могут связаться друг с другом, подобно абонентам, разговаривающими по телефону. Организации такой связи может предшествовать обычный телефонный разговор между владельцами ПК, оговаривающими время связи, скорость передачи информации и прочие параметры протокола. После установления связи между компьютерами по телефонному каналу можно передавать файлы точно так же, как и по «прямому проводу». К услугам подобной связи прибегают абоненты электронной доски объявлений BBS (Bulletin Board System). К существенным недостаткам такого общения между компьютерами относятся непредвиденная занятость телефонной линии, низкая пропускная способность каналов связи, высокая плата за междугородные контакты, необходимость ручного управления сеансом связи со стороны владельца принимающего ПК (клиента), сложность предварительной оценки полезности информации, извлекаемой из сервера BBS. В начале 70-х годов появилась передача данных на основе протокола Internet - IP. Протокол IP - это правила и описание работы сети. Этот свод включает правила налаживания и поддержания связи в сети, правила обращения с IP-пакетами и их обработки, описания сетевых пакетов семейства IP (их структура и т.п.). Сеть задумывалась и проектировалась так, чтобы от пользователей не требовалось никакой информации о конкретной структуре сети. Для того чтобы послать сообщение по сети, компьютер должен поместить данные в некий "конверт", называемый, например, IP, указать на этом "конверте" конкретный адрес в сети и передать получившиеся в результате этих процедур пакеты в сеть. Развитие сферы сетевых услуг привело к массовому появлению серверов, ориентированных на предоставлении информации по определённой тематике. Появились, например, информационные центры

в крупных зарубежных библиотеках, которые перевели большую часть своих фондов в электронный формат и продолжают оперативно вносить в эти архивы новые поступления. Это особенно ценно для наших читателей, т.к. многие отечественные библиотеки влачат жалкое существование. Но библиотечное дело - только капля в информационном океане. Перечень услуг, предоставляемых серверами Internet, достигает нескольких тысяч наименований.**4.2**

## **Конец ознакомительного фрагмента.**

Текст предоставлен ООО «Литрес».

Прочитайте эту книгу целиком, [купив полную легальную версию](#) на Литрес.

Безопасно оплатить книгу можно банковской картой Visa, MasterCard, Maestro, со счета мобильного телефона, с платежного терминала, в салоне МТС или Связной, через PayPal, WebMoney, Яндекс.Деньги, QIWI Кошелек, бонусными картами или другим удобным Вам способом.