

Текст и гипертекст

Лекция №6





Текстовые файлы

Текстовый файл - компьютерный файл, содержащий текстовые данные, как правило, организованные в виде строк.

Текстовый файл, как и прочие файлы, хранится в файловой системе. Текстовым файлам противопоставляются двоичные файлы, в которых содержатся данные, не рассчитанные на интерпретацию в качестве текста (например, файлы, хранящие закодированные звук или изображение).



Текстовый формат

В отличие от термина «**текстовый формат**», характеризующего содержимое данных, термин «**текстовый файл**» относится к самому файлу как контейнеру, хранящему такие данные.

Текстовый файл может содержать не только текст как таковой, но и текст с разметкой в виде текстовых пометок - к примеру, HTML-текст.



Текстовый файл

Текстовый файл представляет из себя последовательность символов и знаков препинания (принадлежащих тому или иному набору символов).

Эти символы обычно сгруппированы в строки (в современных системах строки разделяются разделителями строк, хотя в прошлом применялось хранение строк в виде записей постоянной или переменной длины).

Иногда конец текстового файла (особенно если в файловой системе не хранится информация о размере файла) также отмечается одним или более специальными знаками, известными как маркеры конца файла.

Текстовый файл может содержать как форматированный, так и неформатированный текст.



Текстовые файлы

Преимущества текстовых форматов:

Минимальный объём файла при малом количестве текстовых данных.

Универсальность - текстовый файл может быть прочитан на любой системе или ОС, особенно если речь идёт об однобайтных кодировках вроде ASCII - они не подвержены многим проблемам, характерным для других форматов файлов - для них не важна разница в порядке байтов или длине машинного слова на разных платформах.



Текстовые файлы

Устойчивость - каждое слово и символ в таком файле самодостаточно, и если случится повреждение такого файла, то в этом случае обычно легче восстановить данные и продолжить обработку остального содержимого - повреждение любого из байтов такого файла никак не сказывается на сохранности остальных, в то время как у сжатых или двоичных (например, ***.docx**) иногда при повреждении даже нескольких байтов файл восстановлению не подлежит.



Текстовые файлы

Многие распространённые системы управления версиями рассчитаны на текстовые файлы, и с двоичными файлами могут работать только как с единым целым.



Текстовые файлы

Недостатки текстовых форматов:

Недостатком больших (и несжатых) текстовых файлов является их низкая информационная энтропия - эти файлы занимают больше места, нежели минимально необходимо.

Однако это и определяет высокую устойчивость к сбоям в каналах передачи данных и при воспроизведении данных с носителей, например с магнитной ленты.



Текстовые файлы

Недостатки текстовых форматов:

Эта устойчивость почти не автоматизируема: некоторые ошибки можно находить автоматически, но восстанавливать текст приходится вручную.

Некоторые операции с текстовыми файлами чрезвычайно неэффективны. Например, если в файле встретится число, машина должна будет перевести его в свой внутренний формат, вызвав (сравнительно) сложную процедуру конвертации числа.



Текстовые файлы

Недостатки текстовых форматов:

Чтобы перейти на 1000-ю строку, требуется считать 999 строк, идущих до неё.

Сложно заменить одну строку другой, и т. д.



Гипертекст

Гипертекст - термин, введённый Тедом Нельсоном в 1965 году для обозначения «текста ветвящегося или выполняющего действия по запросу».

Обычно гипертекст представляется набором текстов, содержащих узлы перехода между ними, которые позволяют избирать читаемые сведения или последовательность чтения.



Гипертекст

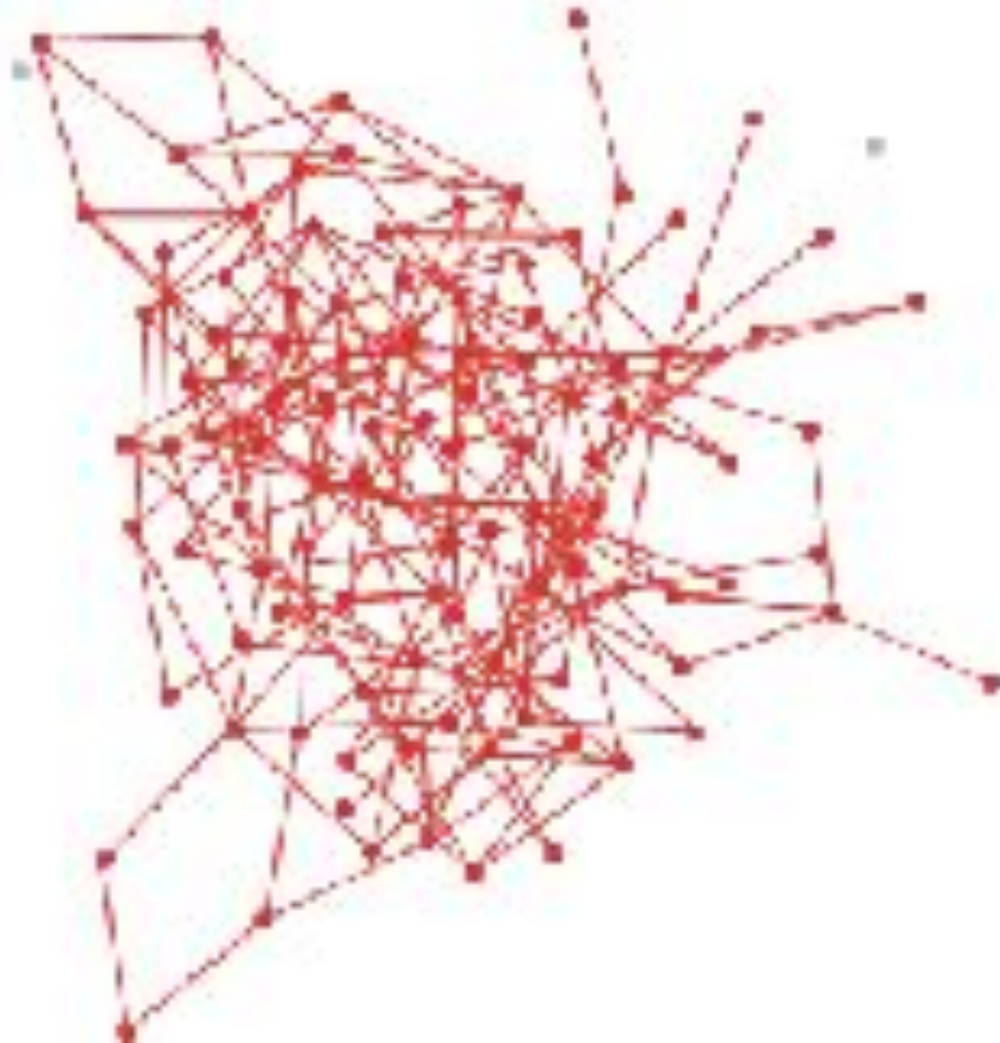
Общеизвестным и ярко выраженным примером гипертекста служат веб-страницы - документы HTML, размещённые в Сети.

В более широком понимании термина, гипертекстом является любая повесть, словарь или энциклопедия, где встречаются отсылки к другим частям данного текста, имеющие отношения к данному термину.

В компьютерной терминологии, гипертекст - текст, сформированный с помощью языка разметки, потенциально содержащий в себе гиперссылки.



Гипертекст как сеть





Гипертекст в литературе

Показателен пример – Библия - Книга Книг сделана как гипертекст со множеством взаимных отсылок к различным главам и строкам документа, именно в связи с тем, что многие происходящие события описываются сразу группой авторов...

Мишель Фуко использует метафору муравейника при обсуждении взаимного влияния многочисленных текстов друг на друга. *Какой статус следует закрепить за дневниками, заметками, записями слушателей, короче говоря, за всем тем муравейником словесных следов, который человек оставляет после смерти и которые обретают голос в бесконечном пространстве языков?* (Фуко М., Археология знания, Киев, 1996, С. 25).



История развития гипертекста

Ваннавер Буш

Впервые устройство, которое бы поддерживало гипертекстовые возможности чтения и письма описал Ваннавер Буш, назвав эту систему Memex.

Прообраз гипертекстового устройства – система Memex, которую Ваннавер Буш описал в статье "Как бы могли мыслить", по своей сути, представляла систему для обмена «мемами».

Гипертекст изначально мыслился создателям как система общественной деятельности.



История развития гипертекста

Группа взаимосвязанных сообщений образовывала сеть, и эта гипертекстовая сеть документов поддерживала социальную сеть отношений между сообществом авторов коллективного гипертекста.

Гипертекстом называют любой текст, отдельные фрагменты, записи или страницы которого обнаруживаются какие-либо ссылки на другие записи.

Используя старую терминологию можно сказать, что гипертекст это - комплекс связанных между собой мемов.



История развития гипертекста

С технической точки зрения гипертекстовая система это - информационная система, способная хранить информацию в виде электронного текста, позволяющая устанавливать электронные связи между любыми "информационными единицами", хранящимися в ее памяти.

Специальные механизмы и правила позволяют компьютеру поддерживать ссылки из одних текстовых фрагментов в другие. Человек или программный агент может устанавливать новые связи между текстовыми фрагментами.



История развития гипертекста

Дуглас Энгельбарт

Вторая после Буша по значимости личность в истории гипертекста это - Даглас Энгельбарт, работавший над проектом расширения мыслительных возможностей человека.

Энгельбарт и его группа сосредоточили свои усилия на обеспечении и расширении познавательных возможностей группы людей.

Многие из возможностей нашли свое широкое применение относительно недавно.



История развития гипертекста

К этим возможностям относятся встроенные в систему возможности установления гипертекстовых связей; возможность хранения групповых переговоров, встроенная в системы электронной почты; возможности личных настроек, перестроек и расширений системы пользователями; возможности усиления не только индивидуальных, но и групповых возможностей.

Файлы в NLS содержались как иерархии сегментов. Каждый сегментов назывался "утверждением".



История развития гипертекста

Каждое "утверждение" снабжалось идентификатором своего уровня в иерархической структуре файла.

Можно было установить любое число ссылочных связей "утверждений" друг с другом, связей как внутрифайловых, так и межфайловых.

В результате структура приобретала неиерархические и нелинейные свойства.

В системе обеспечивалось несколько способов перемещения внутри файла по "утверждениям".



История развития гипертекста

Исследования, проводившиеся в рамках создания системы NLS расширяли возможности сохранения записей в коллективной памяти и, самое главное, заметно упрощали механизмы обмена записями или мимами внутри сетевого сообщества



Области применения компьютерной графики

Нельсон Теодор

Теодор Нельсон, последователь Энгельбарта, работал над созданием всеобщей системы электронных публикаций и всеобщего архива.

Нельсон предложил сам термин "гипертекст" и многое сделал для того, чтобы идеи гипертекста получили широкое распространение

Он неоднократно подчеркивал, что гипертекст в его понимании не является иерархической структурой.



Области применения компьютерной графики

С его точки зрения, знак равенства, который ставят между понятиями «иерархия» и «структура», представляет собой популярный миф.

Живые формы информационных структур не могут быть верно представлены иерархией.

Такие структуры как параллелизм, перекрестные связи, взаимное проникновение и одновременное присутствие одного элемента в нескольких местах не могут быть переданы иерархией.

Гипертекст видится Нельсону как мультиагентное сообщество, внутри которого существуют сложные неиерархические отношения между агентами.



Области применения компьютерной графики

Бернерс-Ли

Бернерс-Ли - отец Всемирной паутины.

Все предыдущие предложения основывались на единой централизованной базе данных - это позволило бы избежать ошибок при переходе от документа к документу.

Бернерс-Ли решил, что для истинной глобальности Сети нужно рассредоточить базу данных по всему миру, сознательно рискуя при этом получить разрывы связей, а ее основой должна стать Сеть сетей - Интернет.



Области применения компьютерной графики

В начале 90-х годов Тим Бернерс-Ли разработал протокол HTTP, который позволил связать между собой документы, размещенные на одном или на нескольких компьютерах, подключенных к сети Интернет.

Эта **гипертекстовая** сеть документов известна как Всемирная Паутина.

Всемирная паутина, как и большинство компьютерных средств, расширяет не только возможности индивидуума, но и возможности группы, сообщества и общества, обеспечивая группе возможное поле для совместной интеллектуальной деятельности.



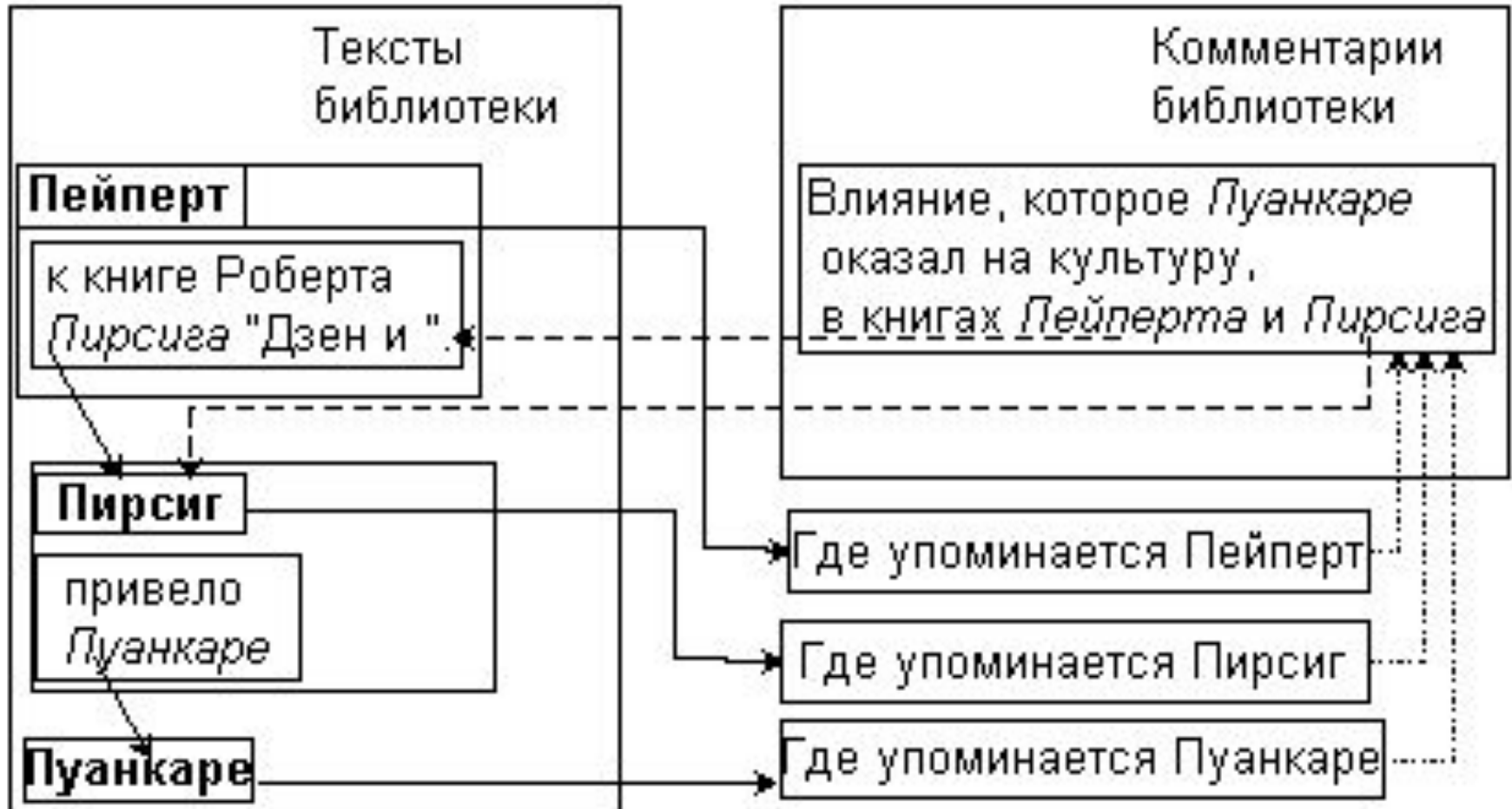
Области применения компьютерной графики

Сам Бернерс-Ли описывал Всемирную паутину как абстрактное пространство информации. Паутина делает сеть полезной, поскольку люди на самом деле интересуются информацией и не хотят ничего знать про провода и компьютеры. Паутина существует, потому что программы работают и поддерживают постоянный обмен информацией между компьютерами.

В начале работ по созданию Всемирной Паутины Бернерс-Ли представлял ее как общее информационное пространство, в котором люди общаются и делятся информацией.



Схема гипертекста





Области применения компьютерной графики

Гипертекстом называют любой текст, в котором обнаруживаются какие-либо ссылки на другие фрагменты. Гипертекстовая система это - информационная система, способная хранить информацию в виде электронного текста, позволяющая устанавливать электронные связи между любыми "информационными единицами", хранящимися в ее памяти и вызывать их на экран монитора "простым нажатием кнопки".

Концепция гипертекста достаточно проста. Есть база данных, в базе данных находятся объекты. Объекты это, чаще всего, небольшие текстовые разделы, посвященные тому или иному вопросу. Специальные механизмы и правила позволяют компьютеру поддерживать ссылки из одних текстовых фрагментов в другие. Человек или программный агент может устанавливать новые связи между текстовыми фрагментами. Система текстовых фрагментов или файлов с такой организацией получила название "гипертекст".



Области применения компьютерной графики

- ✓ **Специфика гипертекста**
- ✓ **Специфика гипертекста как определенной модели текста заключается как в нелинейности его построения, так и восприятия. К основным композиционным принципам построения гипертекста относят:**
 - ✓ **1. дискретность подачи информации, то есть отсутствие непрерывности;**
 - ✓ **2. нелинейность;**
 - ✓ **3. антропоцентричность, или субъективизм в отношении коммуникативной завершенности гипертекста;**
 - ✓ **4. «демократичность» - отдельные части гипертекста, в зависимости от позиции, избранной читателем, могут служить и началом, и завершением одновременно.**
- ✓ **Часто понятие гипертекста связывают с понятиями интертекстуальности и интертекста. Под интертекстуальностью понимают многомерную связь отдельного текста с другими прецедентными текстами - по линиям содержания, жанрово-стилистическим особенностям, структуре, формально-знаковому выражению.**
- ✓ **Под интертекстом понимается основной вид и способ построения художественного текста, состоящий в том, что текст строится из цитат, реминисценций и аллюзий к другим текстам. Интертекст и гипертекст представляют собой системы текстов, отличающиеся друг от друга выраженностью связей между текстами. И интертекст, и гипертекст содержат ссылку на другой текст, но в первом случае эта ссылка присутствует имплицитно, а во втором - эксплицитно. Ссылка на своего предшественника в интертексте неразрывно связана с канвой всего повествования в целом, т.е. не может быть обойдена читательским вниманием. Ссылка в гипертексте не предполагает наличия прецедентного текста, она более «демократична», адресат может обратиться к ней, а может и проигнорировать.**
- ✓ **Особенность гипертекста как структуры заключается в способности, сохраняя сетевой характер, переходить в форму упорядоченной по смыслу последовательности высказываний, в форму текста. В то же время гипертекст открыт для новых элементов и траекторий, которые не нарушают, а лишь обогащают его структуру. В гипертексте субъект самостоятельно движется по цепочке гипертекстовых связей, создавая свою информационную модель мироздания, создавая свой смысл данной модели.**