

ПОШУК, ОБРОБКА ТА АНАЛІЗ ІНФОРМАЦІЇ

СТОРЧАК К.П., ТКАЛЕНКО О.М.
ПОЛОНЕВИЧ О.В., КОСЕНКО В.Р., ЧОРНА В.М.

ПОШУК, ОБРОБКА ТА АНАЛІЗ ІНФОРМАЦІЇ



НАВЧАЛЬНИЙ ПОСІБНИК

Київ
ДУТ
2018

УДК 621.395.57
681.334
ББК 32.895

Розглянуто та затверджено на засіданні кафедр
Інформаційних систем та технологій
Протокол №10 від 30 квітня 2018 року

Навчальний посібник призначений для студентів денної та заочної форми навчання за всіма спеціальностями університету, а також може бути корисним для аспірантів, викладачів навчальних закладів відповідних спеціальностей.

Сторчак К.П., Ткаленко О.М., Полоневич О.В., Косенко В.Р., Чорна В.М.
Пошук, обробка та аналіз інформації. Навч.посібник, підготовлено для студентів вищих навчальних закладів – Київ: ДУТ, 2018.-127 с.

Рецензенти:

Заїка В.Ф. – Завідувач кафедрою Телекомунікаційних систем та мереж Державного університету телекомунікацій, доктор технічних наук, доцент.
Кунах Н.І. – доктор технічних наук, професор Київського коледжу зв'язку.

ЗМІСТ

ВСТУП.....	4
1 СТРАТЕГІЯ ПРОФЕСІЙНОГО ІНФОРМАЦІЙНОГО ПОШУКУ..	5
1.1 Основи теорії інформації.....	5
1.1.1 Визначення поняття «інформація». Її основні властивості...	5
1.1.2 Основні інформаційні процеси.....	9
1.2 Пошук інформації в мережі Інтернет.....	13
1.3 Критерії оцінки якості пошуку.....	17
2 ПОШУКОВІ СИСТЕМИ ТА КОРИСТУВАННЯ НИМИ.....	20
2.1 Модель Web-простору.....	20
2.2 Принцип роботи пошукових машин.....	21
2.3 Ранжування інформації пошуковими системами.....	28
2.4 Найбільш популярні пошукові системи світу та їх особливості.....	31
2.5 Мета-пошукові системи.....	39
2.6 Основи роботи з пошуковою системою Google.....	40
2.6.1 Особливості пошуку в Google.....	40
2.6.2 Найбільш вживані оператори пошуку.....	44
2.6.3 Найбільш популярні спеціальні оператори Google.....	48
2.6.4 Як шукати за зображенням.....	49
3 ОБРОБКА ІНФОРМАЦІЇ.....	50
3.1 Кодування інформації.....	51
3.2 Засоби обробки інформації.....	55
3.3 Технології обробки інформації.....	62
3.3.1 Основна мета та історія розвитку.....	62
3.3.2 Інформаційні технології в умовах функціонування систем обробки інформації.....	66
3.3.3 Основні режими обробки інформації.....	68
3.3.4 Комплекс технічних засобів обробки інформації.....	73
4 ЗБЕРІГАННЯ ІНФОРМАЦІЇ.....	77
4.1 Збір та зберігання інформації.....	77
4.2 Зберігання інформації в персональному комп'ютері.....	81
4.3 Оперативна пам'ять.....	89
4.4 Зберігання інформації в Інтернеті.....	100
4.5 Основні системи зберігання даних та їх особливості.....	106
СПИСОК ЛІТЕРАТУРИ.....	127

ВСТУП

Сьогодні в умовах інформаційного суспільства вміння грамотно здійснювати професійний пошук, аналіз і обробку інформації - незаперечний навик фахівця-інтелектуала.

Швидке зростання і динамічний розвиток інформаційних ресурсів (перш за все представлених в мережі Інтернет) зобов'язує кожного мати різноманітні навички складного пошуку. Вміти швидко і правильно шукати в Інтернеті означає економити час, володіти достовірною та актуальною інформацією, а значить робити правильні висновки і приймати правильні рішення.

Інформаційна евристика покликана допомогти фахівцеві досить швидко провести виборчий і ефективний пошук при вирішенні конкретних інтелектуальних завдань.

Проблема пошуку і використання інформації - одна з найактуальніших в сучасній науці. Зрозуміло, для ефективного і якісно пошуку інформації Інтернет незамінний. Однак сам факт зберігання в Інтернеті величезного пласту знань людства не є свідченням високої ефективності пошуку і використання інформації. Нехай навіть інформація видається і в зручній формі, але освоювати її в кінцевому підсумку доводиться людині. І від того, наскільки він володіє методами і методикою пошуку, буде залежати ефективність і якість його праці.

Отже, в основі використання інформації лежить проблема пошуку і перетворення її в такий зміст і форму, які створюють можливість більш зручного і оперативного освоєння інформації та її ефективного використання в громадській діяльності.

1 СТРАТЕГІЯ ПРОФЕСІЙНОГО ІНФОРМАЦІЙНОГО ПОШУКУ

1.1. Основи теорії інформації.

1.1.1 Визначення поняття «інформація». Її основні властивості.

Слово «інформація» походить від латинського слова *informatio*, що в перекладі означає зведення, роз'яснення, ознайомлення. В середині 20 століття термін «інформація» став загальнонауковим поняттям, що включає обмін відомостями між людьми, людиною і автоматом (ЕОМ), автоматом і автоматом, обмін сигналами в тваринному і рослинному світі, передачу ознак від клітини до клітини, від організму до організму.

Інформація - сукупність відомостей про об'єкти, явища, процеси, їх властивості і відносини. [1]

Повідомлення - сукупність символів кінцевого алфавіту, що є формою вираження інформації. [1]

Сигнал - матеріальний носій інформації, що володіє змінними параметрами (звук, світло, радіо сигнали, напруга, кутове або лінійне переміщення і т.д.). [1]

Дані (англ. Data) - це інформація зібрана і трансформована для деяких цілей, зазвичай аналізу.

Інформацію можна розділити на види за різними критеріями [4]:

1. За способом сприйняття:

- візуальна - сприйнята органами зору;
- звукова - сприйнята органами слуху;
- тактильна - сприйнята тактильними рецепторами;
- нюхова - сприйнята нюховими рецепторами;
- смакова - сприйнята смаковими рецепторами.

2. За формою подання:

- текстова - що передається у вигляді символів, призначених позначати лексеми мови;

- числова - у вигляді цифр і знаків, що позначають математичні дії;
- графічна - у вигляді зображень, предметів, графіків;
- звукова - усна або у вигляді запису і передачі лексем мови аудіальним шляхом;

- відеоінформація - передана у вигляді відеозапису.

3. За призначенням:

- масова - містить тривіальні відомості і оперує набором понять, зрозумілим більшій частині соціуму;

- спеціальна - містить специфічний набір понять, при використанні відбувається передача відомостей, які можуть бути не зрозумілі основній масі соціуму, але необхідні і зрозумілі в рамках вузької соціальної групи, де використовується дана інформація;

- секретна - передана вузькому колу осіб і за закритими (захищеними) каналами;

- особиста (приватна) - набір відомостей про які-небудь особистості, що визначає соціальний стан і типи соціальних взаємодій всередині популяції.

Основні властивості інформації:

1. Релевантність – здатність інформації відповідати потребам (запитам) споживача;

2. Повнота – властивість інформації вичерпно (для даного споживача) характеризувати об'єкт та/або процес, що відображається;

3. Своєчасність – здатність інформації відповідати потребам споживача в потрібний момент часу;

4. Достовірність – властивість інформації не мати прихованих помилок;

5. Доступність – властивість інформації, що характеризує можливість її отримання даними споживачем;

6. Захищеність – властивість, що характеризує неможливість несанкціонованого використання або зміни;

7. Ергономічність – властивість, що характеризує зручність форми або обсягу інформації з точки зору даного споживача;

8. Адекватність – властивість інформації однозначно відповідати об'єкту або явищу, що відображається. Адекватність для споживача є внутрішньою властивістю інформації, що проявляється через релевантність і достовірність;

9. Живучість – здатність інформації зберігати свою якість з плином часу;

10. Унікальною називають інформацію, що зберігається в єдиному екземплярі.

Живучість і унікальність є властивостями інформації, які пов'язані з процесом її зберігання.

Закони існування інформації.

1. Загальний закон збереження інформації.

Інформація в загальному вигляді, будучи властивістю всіх матеріальних об'єктів, існує вічно, ніколи не виникала і ніколи не зникає. Ми бачимо стіл - інформацію про нього ми отримуємо як тільки його побачили, але якщо ми його не бачимо це не означає, що інформації про нього немає - просто вона нам не доступна.

2. Загальний закон переходу інформації.

Усе інформаційне поле обумовлено взаємодіями структурної та оперативної інформації та їх взаємними переходами (перетвореннями) одна в одну.

Структурна (або зв'язна), притаманна об'єктам неживої природи природного і штучного походження, наприклад стіл - ми, бачачи або сприймаючи його, отримуємо інформацію про ньому - матеріал, форма, колір і т.п.; ця інформація невід'ємна від об'єкта (стола), вона притаманна саме цьому столу.

Оперативна (або робоча), це інформація що циркулює між об'єктами матеріального світу і використовується в процесах управління в живій природі, в людському суспільстві, це в першу чергу команди (явні і приховані) і інші форми управління. Вся неструктурна інформація є оперативною.

Всі інші закони є приватними по відношенню до двох вищеназваних і відносяться до функціонування оперативної інформації.

3. Закон незбереження інформації.

При будь-яких змінах системи-носія інформації, що призводять до порушення структури повідомлення (або відображення), інформація зникає в ній частково або повністю у відповідності з розмірами порушень. При цьому вона не перетворюється ні в речовину, ні в енергію, ні в іншу інформацію, а просто зникає.

Людина бачив дерево влітку. За осені дерево скинуло листя - змінилося. Той же людина бачить це ж дерево але інформація вже інша.

4. Закон узгоджених каналів інформаційної взаємодії.

Дві системи (об'єкта) можуть здійснювати інформаційну взаємодію тільки через узгоджені канали зв'язку. Люди, розмовляючи, видають звуки в одному звуковому діапазоні. В іншому випадку вони не почують один одного. Радіостанції працюють на одній частоті - інакше вони не почують одна одну.

5. Закон подвоєння інформації.

В процесі одиничного акту передачі інформації від одного об'єкта до іншого за погодженим каналом зв'язку, вона (інформація) подвоюється - другий об'єкт в цьому випадку інформацію отримує, а перший її не втрачає. Наприклад повідомлення від людини людині - той хто повідомлення передав продовжать залишатися власником цієї інформації, і той хто повідомлення прийняв так само стає володарем цієї інформації. У матеріальному світі все трохи інакше - якщо ти передав предмет іншому, то втратив його для себе.

6. Закон перекодування інформації.

Передача інформації від одного об'єкта до іншого з будь-якого погодженим каналом зв'язку здійснюється лише шляхом перекодування; пряма передача інформації в цьому випадку неможлива. Ми не вміємо здійснювати прямий обмін інформацією - ми модулюємо цією інформацією якісь підвладні нам канали зв'язку і таким чином обмінюємося інформацією. Це мова, писемність, знаки, відео, зображення.

7. Закон тезауруса.

Для адекватного сприйняття отриманого об'єктом повідомлення, останній повинен мати в наявності апіорну інформацію (тезаурус), достатню для дешифрування і засвоєння отриманого повідомлення. Для мовлення це алфавіт, для мови - словник, для управління - набір команд.

8. Закон фасцинації.

Для адекватного сприйняття отриманого об'єктом повідомлення, останнє повинно мати привабливу форму. Кардинальний варіант - відторгнення інформації яка має негативну форму, форму неприємну для приймаючої сторони.

9. Закон маєвтики.

Інформація, яка надходить об'єкту, може бути розглянута не тільки як корисне повідомлення, але і як стимул для породження нової інформації на основі минулого досвіду і моделювання ситуації. Слідство: в результаті маніпуляцій інформацією в спеціальних (логічних, інтуїційних, кібернетичних) системах її переробки може бути отримана нова інформація без додаткового надходження ззовні. (Термін "маєвтика" в перекладі з грецької мови означає "породіллі" і сходить ще до Сократа.) А це власне робота аналітика - отримувати інформацію, переробляти її і видавати нову інформацію.

10. Закон опосередкованого управління.

Інформація, може бути опосередковано використана як "агент" управління певним об'єктом. Наприклад керуючий вплив людини на людину.

1.1.2. Основні інформаційні процеси

Можливо виділити п'ять основних інформаційних процесів [2].

1. Збір інформації.

Один з основних процесів є збір інформації.

Процес збору інформації являє собою діяльність суб'єкта, метою якої є отримання відомостей про об'єкт, який його цікавить.

Збір інформації може здійснюватися або людиною, або за допомогою технічних засобів і систем – апаратно. Система збору інформації може являти собою складний програмно-апаратний комплекс. Як правило, сучасні системи збору електронної інформації не тільки забезпечують кодування інформації і її введення в комп'ютер, а й виконують попередню (первинну) обробку цих даних.

2. Обробка інформації.

Схема обробки інформації: вихідна інформація - виконавець обробки - підсумкова інформація.

У процесі обробки інформації вирішується деяка інформаційна задача, яка попередньо може бути поставлена в традиційній формі: даний деякий набір вихідних даних, потрібно отримати деякі результати. Сам процес переходу від вихідних даних до результату і є процес обробки. Об'єкт або суб'єкт, який здійснює обробку, називають виконавцем обробки.

Для успішного виконання обробки інформації виконавцю (людині або пристрою) повинен бути відомий алгоритм обробки, тобто послідовність дій, яку потрібно виконати, щоб досягти потрібного результату.

Розрізняють два типи обробки інформації:

Перший тип обробки: обробка, пов'язана з отриманням нової інформації, нового змісту знань (рішення математичних задач, аналіз ситуації та ін.). Другий тип обробки: обробка, пов'язана зі зміною форми, але не змінює змісту (наприклад, переклад тексту з однієї мови на іншу).

Важливим видом обробки інформації є кодування - перетворення інформації в символну форму, зручну для її зберігання, передачі, обробки. Кодування активно використовується в технічних засобах роботи з інформацією (телефонія, комп'ютери).

Інший вид обробки інформації - структурування даних (внесення певного порядку в сховище інформації, класифікація, каталогізація даних).

3. Передача інформації.

Розвиток людства був би неможливим без обміну інформацією. З давніх часів люди з покоління в покоління передавали свої знання, сповіщали про небезпеку або передавали важливу і термінову інформацію, обмінювалися відомостями.

У будь-якому процесі передачі або обміні інформацією існує її джерело і одержувач, а сама інформація передається по каналу зв'язку за допомогою сигналів: механічних, теплових, електричних і ін.

Інформація представляється і передається в формі послідовності сигналів, символів. Від джерела до приймача повідомлення передається через деяке матеріальне середовище.

Як джерело інформації може виступати жива істота або технічний пристрій. Від нього інформація потрапляє в кодуєчий пристрій, який призначений для перетворення вихідного повідомлення в форму, зручну для передачі. З такими пристроями ви зустрічаєтеся постійно: мікрофон телефону, лист паперу і т.д.

Каналом зв'язку інформація потрапляє в декодер одержувача, яке перетворює кодоване повідомлення у форму, зрозумілу одержувачу. Одні з найскладніших декодер - людське вухо та око.

В процесі передачі інформація може втрачатися, спотворюватися. Це відбувається через різні перешкоди, як в каналі зв'язку, так і при кодуванні і декодуванні інформації.

Терміном «шум» називають різного роду перешкоди, які спотворюють сигнал, що передається і призводять до втрати інформації. Такі перешкоди, перш за все, виникають з технічних причин: погана якість ліній зв'язку, незахищеність один від одного різних потоків інформації, переданої по одним і тим же каналам. Для захисту від шуму застосовуються різні способи, наприклад, застосування різного роду фільтрів, що відокремлюють корисний сигнал від шуму.

4. Зберігання.

Збір інформації не є самоціллю. Щоб отримана інформація могла використовуватися, причому багаторазово, необхідно її зберігати.

Зберігання інформації - це спосіб поширення інформації в просторі і часі.

Людський розум є найдосконалішим інструментом пізнання навколишнього світу. А пам'ять людини - прекрасним пристроєм для зберігання отриманої інформації.

Щоб інформація стала надбанням багатьох людей, необхідно мати можливість її зберігати не тільки в пам'яті людини. У процесі розвитку людства існували різні способи зберігання інформації, які удосконалювалися з часом: вузлики на вірьовках, карби на палицях, берестяні грамоти, листи на папірусі, папері.

Нарешті, був винайдений друкарський верстат, і з'явилися книги. Пошук надійних і доступних способів зберігання інформації йде і до цього дня.

Сьогодні ми використовуємо для зберігання інформації найрізноманітніші матеріали: папір, фото- і кіноплівку, магнітні аудіо- та відеострічки, магнітні та оптичні диски. Все це - носії інформації.

Носій інформації - матеріальний об'єкт, призначений для зберігання і передачі інформації.

Сховище інформації - це певним чином організована інформація на зовнішніх носіях, призначена для тривалого зберігання і постійного використання (наприклад, архіви документів, бібліотеки, картотеки).

Основною інформаційною одиницею сховища є певний фізичний документ: анкета, книга та ін. Під організацією сховища розуміється наявність певної структури, тобто впорядкованість, класифікація збережених документів для зручності роботи з ними.

Основні властивості сховища інформації: обсяг інформації, що зберігається, надійність зберігання, час доступу (тобто час пошуку потрібних відомостей), наявність захисту інформації.

Інформацію, збережену на пристроях комп'ютерної пам'яті, прийнято називати даними. Організовані сховища даних на пристроях зовнішньої пам'яті комп'ютера прийнято називати базами і банками даних.

5. Пошук.

Інформаційним пошуком (ІП) називається деяка послідовність операцій, що виконуються з метою відшукування документів, що містять певну інформацію, або з метою видачі фактичних даних, що представляють собою відповіді на певні питання.

Виділяють наступні методи пошуку інформації: безпосереднє спостереження; спілкування з фахівцями з тем, які Вас цікавлять; читання відповідної літератури; перегляд відео, телепрограм; прослуховування радіопередач, аудіокасет; робота в бібліотеках і архівах; запит до інформаційних систем, баз і банків комп'ютерних даних; інші методи.

Для того щоб зібрати як найповнішу інформацію і підвищити ймовірність прийняття правильного рішення, необхідно використовувати різноманітні методи пошуку інформації.

1.2 Пошук інформації в мережі Інтернет

Пошук інформації в мережі Internet - це послідовність дій, від визначення предмету пошуку, до отримання відповіді на наявні питання з використанням всіх пошукових сервісів, які надає Інтернет сьогодні. [3]

Основні переваги використання мережі Інтернет при пошуку інформації:

- використання максимально можливого «простору пошуку» інформації. Жоден з існуючих на сьогодні мережевих ресурсів не володіє тим обсягом інформації, який представлений в Інтернеті;
- жодне інше джерело не володіє такою оперативністю і доступністю. Інтернет надає доступ цілодобово незалежно від місця знаходження;
- інформацію, отриману через Інтернет можна легко переслати для обговорення або, наприклад, роздрукувати в потрібному числі примірників.

Основними етапами методики пошуку інформації в мережі Internet є:

1. Формулювання і уточнення інформаційного запиту. На цьому етапі йде цілеспрямоване обмеження і конкретизація загальної мети пошуку;
2. Планування пошукової процедури. Основна мета даного процесу - визначення шляхів і способів раціонального рішення пошукового завдання;
3. Реалізація пошуку. Остаточне рішення пошукової завдання.

Перший етап має велике значення та у великій мірі визначає, чи знайдете Ви потрібну Вам інформацію. Тому дуже важливо навчитися підбирати ключові слова і їх комбінації. На основі попередньо отриманої інформації про предмет пошуку потрібно скласти максимально широкий набір ключових слів у вигляді окремих термінів, словосполучень, назв і прізвищ, тісно пов'язаних з проблемою, професійної лексики.

Природна мова запитів приваблює своєю простотою, однак, як показує практика, він менш ефективний при серйозному пошуку, коли ставиться завдання знайти всі або більшість документів на задану тему.

Основні моменти, які допомагають правильно сформулювати запит до пошукової машини:

- не слід ставити тільки одне слово. Потрібно використовувати цілі фрази, або, хоча б, кілька слів;
- слід вказувати слова, які не повинні зустрічатися в шуканих документах. Зазвичай для цього використовують або знак "-", або ключове слово NOT;
- якщо потрібно знайти фразу цілком – потрібно взяти її в лапки;
- якщо Ви вводите запит до пошукової машини, що складається з декількох слів, то в результаті отримуєте список документів, в яких зустрічається хоча б одне слово.

Більш детально нюанси формування пошукових запитів описано у другому розділі даного навчального посібника.

Другий етап – планування пошукової процедури. Можна виділити наступні основні методи пошуку інформації в Інтернеті, які, в залежності від

цілей і завдань шукача, використовуються окремо або в комбінації один з одним:

1. через вказівку прямої адреси сайту;
2. безпосередній пошук з використанням гіпертекстових посилань;
3. використання пошукових систем;
4. пошук із застосуванням спеціальних засобів;
5. аналіз нових ресурсів.

Через вказівку прямої адреси сайту.

Цей метод передбачає введення прямої адреси сайту, де розміщені потрібні дані і необхідна інформація. Це єдиний спосіб знаходити «острови» в Інтернеті.

Безпосередній пошук з використанням гіпертекстових посилань.

Оскільки всі сайти в просторі WWW фактично виявляються пов'язаними між собою, пошук інформації може бути здійснений шляхом послідовного перегляду пов'язаних сторінок за допомогою браузера.

Інтернет серфінг - пошук, при якому здійснюється послідовний перехід по посиланнях з одного тематичного сайту на Інший до тих пір, поки не буде знайдено джерело потрібних Даних.

Перегляд Web-сторінок часто виявляється єдино можливим на заключних етапах інформаційного пошуку.

Використання пошукових систем.

Сьогодні цей метод є одним з основних і фактично єдиним при проведенні попереднього пошуку.

За принципом роботи пошукові системи бувають двох видів: пошукові індекси и пошукові каталоги.

Пошукові каталоги - ці сайти надають допомогу в пошуку даних певної тематики в мережі Інтернет: інформація в таких каталогах чітко структурована за групами і темам, що сприяє швидкому знаходженню результату. Інформація заноситься людиною вручну. Приклад: Rambler, Yahoo!

Пошукові індекси - це сайти-покажчики, в яких при введенні ключового слова в рядок пошуку, користувач отримує ряд посилань на сторінки в інтернеті, де містяться запитовані слово або фраза. Приклад: Google, AltaVista, HotBot, Яндекс та інші.

За допомогою пошукових індексів завжди легше знайти щось конкретне, і важче - щось спільне і невизначене.

Пошук із застосуванням спеціальних засобів

Цей повністю автоматизований метод може виявитися досить ефективним для проведення первинного пошуку. Одна з технологій цього методу заснована на застосуванні спеціалізованих програм - спайдерів, які в автоматичному режимі переглядають Web-сторінки, відшуковуючи на них потрібну інформацію. Фактично це автоматизований варіант перегляду за допомогою гіпертекстових посилань, описаний вище (пошукові машини для побудови своїх індексних таблиць використовують схожі методи). Результати автоматичного пошуку обов'язково вимагають подальшої обробки.

Застосування даного методу доцільно, якщо використання пошукових машин не може дати необхідних результатів (наприклад, в силу нестандартності запиту, який не може бути адекватно заданий існуючими засобами пошукових машин). У ряді випадків цей метод може бути дуже ефективний. Вибір між використанням спайдера або пошукових серверів являє собою варіант класичного вибору між застосуванням універсальних або спеціалізованих засобів.

Аналіз нових ресурсів

Пошук по новоствореним ресурсам може виявитися необхідним при проведенні повторних циклів пошуку, пошуку найбільш свіжої інформації або для аналізу тенденцій розвитку об'єкту дослідження в динаміці. Іншою можливою причиною може бути те, що більшість пошукових машин оновлює свої індекси зі значною затримкою, викликаною гігантськими обсягами оброблюваних даних, і ця затримка зазвичай тим більше, ніж менш популярна

тема, яка Вас цікавить. Це міркування може виявитися досить істотним при проведенні пошуку в вузькоспеціалізованій предметній області.

1.3 Критерії оцінки якості пошуку.

Для того щоб знайдена інформація дійсно відображала реальний стан справ, пошук повинен задовольняти наступним критеріям:

- повнота охоплення ресурсів;
- достовірність інформації;
- висока швидкість проведення пошуку.

Контроль повноти охоплення ресурсів - досить серйозна проблема, в тому випадку якщо необхідно не просто знайти будь-яку інформацію про предмет, який Вас цікавить, а Ви хочете мати повне уявлення про предмет пошуку і стан справ з даного питання. В цьому випадку краще не обмежуватися використанням тільки однієї пошукової машини або переглядом одного, нехай навіть найулюбленішого каталогу. Для того щоб провести повномасштабний збір інформації необхідно працювати з усіма відомими каталогами, пошуковими машинами, базами даних, регіональними телеконференціями, електронними дошками оголошень і листами розсилок. Тільки в цьому випадку Ви будете впевнені, що знайшли якщо не всю, то хоча б більшу частину існуючої інформації. При цьому не варто забувати і про електронні ЗМІ.

При проведенні пошуку інформації в Інтернеті, в якому повнота грає важливу роль, необхідно пам'ятати, що жодна пошукова система не індексує всього, що є в Інтернеті.

Розширити список інформаційних ресурсів можна наступними шляхами:

- провести пошук в зарубіжних мета-пошукових системах з метою виявлення баз даних, присвячених конкретно тематиці пошуку;
- провести пошук по гіпертекстових посиланнях, рухаючись від вже знайдених документів;

- провести пошук баз даних, спираючись на структуру URL вже знайдених документів.

Контроль достовірності інформації. Ще одна важлива і досить складна частина процесу пошуку. В силу своєї специфіки, Інтернет містить достатню кількість застарілою або не достовірної інформації. Багато в чому це пояснюється можливістю анонімного розміщення матеріалів, особливо на безкоштовних сервісах. В основному, контроль достовірності інформації - це аналітична робота. При цьому важливо провести звірку знайденого фактичного матеріалу, з'ясувати статус документів, отримати інформацію про компетентність автора матеріалу і т.д.

Результати пошуку.

Як вже зазначалося вище, при пошуку в Інтернеті важливі:

- повнота охоплення - нічого не втрачено з наявної інформації;
- точність - не знайдено зайвої або недостовірної інформації.

Отримавши потрібну інформацію по мережі, намагайтеся для початку її перевірити. Перевіряється кожна адресна інформація: якими б можливостями не володіла сучасна техніка, але заносять інформацію в мережу люди, а їм властиво помилятися, проявляти неакуратність, та неоперативність і т.д.

Аналітичну інформацію перевіряють, порівнюючи дані, отримані з декількох джерел. Таким же способом перевіряють різноманітні статистичні дані.

Якщо Ви передбачаєте використовувати отриману інформацію для публічного огляду, Ви зобов'язані точно записати джерело (адресу в мережі), ім'я автора (або назву організації) і дату публікації використовуваних відомостей. Зверніть увагу на можливу наявність особливих вказівок щодо дотримання авторських прав.

Швидкість проведення пошуку в мережі, якщо не брати до уваги технічні характеристики підключення користувача, залежить в основному від двох факторів:

- грамотного планування пошукової процедури;

- досвіду роботи з ресурсом вибраного типу.

Особливе значення швидкість проведення пошуку має в тому випадку, коли Ви маєте справу з інформацією, яка швидко оновлюється.

Для того, щоб значно підвищити якість пошуку інформації у Мережі Інтернет, слід скористатися наступними порадами:

1. Щоб знайти правильну адресу сторінки, шукайте одночасно ключові слова із запитання і відомої вам відповіді на поставлене питання;
2. Число документів, отриманих в результаті пошуку, може бути величезна. Тому вирішальне значення для оптимального пошуку інформації має правильний набір ключових слів;
3. Перевіряйте орфографію в написанні слова. Використовуйте синоніми, якщо список знайдених сторінок занадто малий;
4. Шукайте більше, ніж по одному слову. Максимально звужуйте предмет пошуку;
5. Використовуйте налаштування пошуку в пошукових системах, щоб задати зручні параметри опису документів і структуру сторінки з результатами пошуку.

2 ПОШУКОВІ СИСТЕМИ ТА КОРИСТУВАННЯ НИМИ

2.1. Модель Web-простору

Web-простір - Інтернет-простір.

Web-сторінка - документ або інформаційний ресурс Всесвітньої павутини, доступ до якого здійснюється за допомогою веб-браузера.

Дослідження, проведені вченими, спростували поширену думку, ніби Internet - це єдиний густий простір [5]. Простеживши за допомогою пошукового механізму AltaVista понад 200 млн. Web-сторінок і кілька мільярдів посилань, розміщених на цих сторінках, вчені прийшли до висновку про структуру Web-простору як орієнтованого графа, в якому вершини відповідають Web-сторінкам, а ребра - з'єднуючим ці сторінки гіперпосиланням. В рамках цієї моделі завдання аналізу структури зв'язків між окремими Web-сторінками було збудовано наступну схему Web-простору (рис. 2.1).

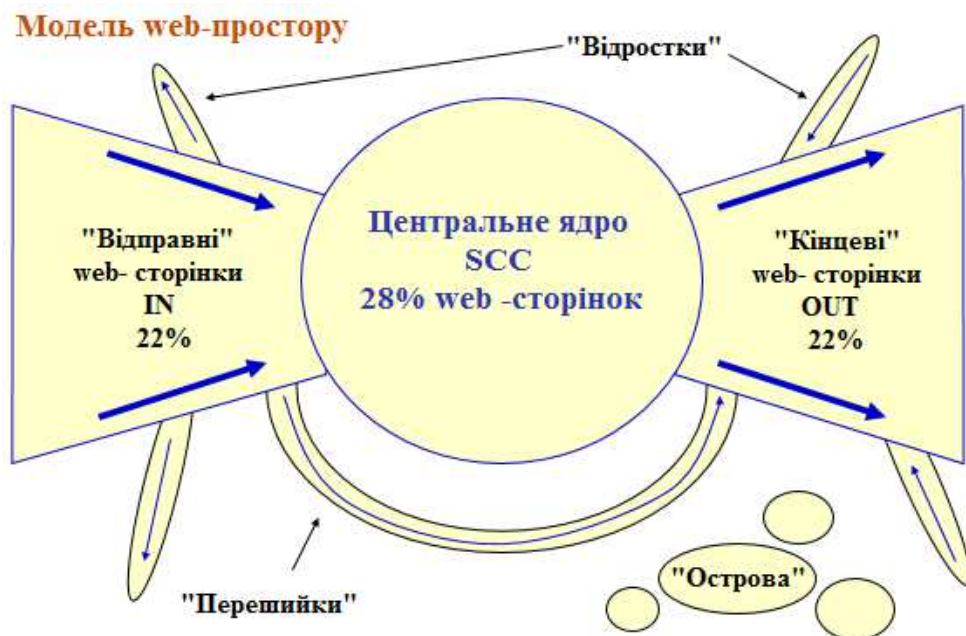


Рисунок 2.1 – Модель web-простору

1. Центральне ядро - 28% web-сторінок.

Компоненти сильної зв'язності (SCC).

Сюди відносяться web-сторінки, пов'язані так тісно, що, слідуючи за гіперпосиланнями, з будь-якої з них в кінцевому рахунку можна потрапити на будь-яку іншу.

2. «Відправні» web-сторінки - 22%.

Web-сторінки, які містять гіперпосилання, що ведуть в кінцевому рахунку до ядра. Але! З ядра за гіперпосиланнями на них потрапити не можна!

3. «Кінцеві» web-сторінки - 22%.

До цих web-сторінок можна прийти по посиланнях з ядра. Але повернутися за гіперпосиланнями назад в ядро з цих сторінок неможливо!

4. «Відростки» - 22%.

Web-сторінки, повністю ізольовані від центрального ядра. Це або «відростки», пов'язані в односторонньому порядку зі сторінками іншої категорії. Або «перешийки», що з'єднують web-сторінки, що не входять в ядро.

5. «Острови» - близько 10%.

Web-сторінки, які взагалі не перетинаються з іншими ресурсами Інтернету. Єдиний спосіб виявити ці сторінки - знати їхню адресу. Ніякі пошукові машини не можуть знайти «острова», якщо на них не ведуть гіперпосилання.

Вчені виявили, що пропорції чотирьох основних категорій web-сторінок протягом часу залишаються незмінними, незважаючи на значне збільшення загального обсягу web-ресурсів.

2.2 Принцип роботи пошукових машин

Пошукова машина (частина пошукової системи) являє собою комплект програм, в основі якого лежать наступні п'ять: Spider («павук»), Crawler («Черв'як», або «мандрівний павук»), Indexer (індексатор), Search Engine Results Engine (система видачі результатів пошуку) (рис. 2.2).

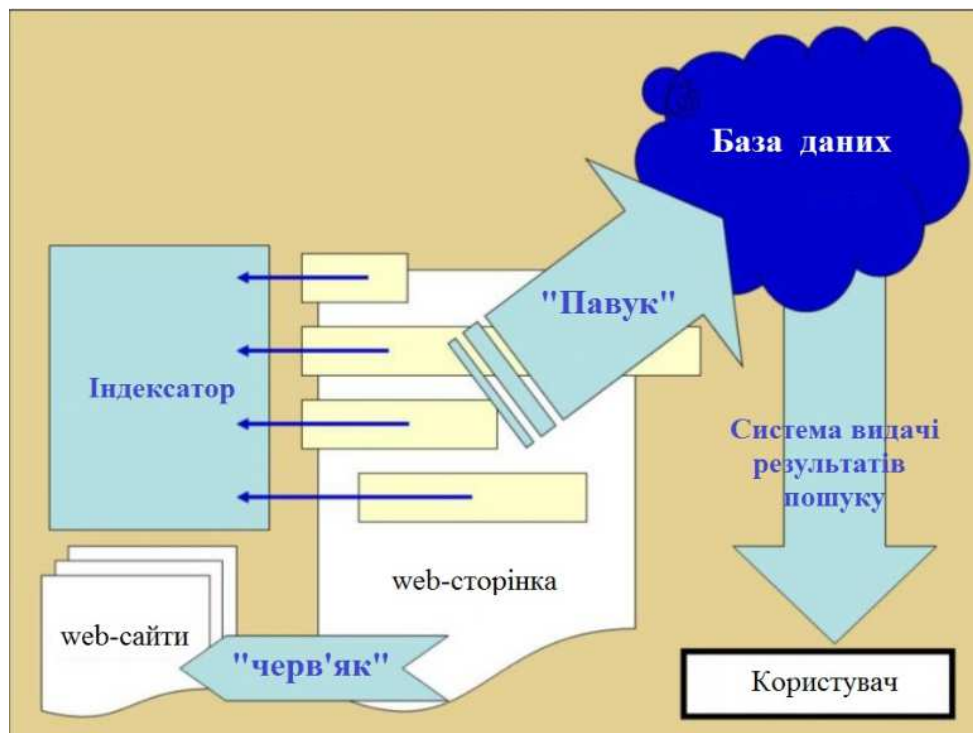


Рисунок 2.2 – Устрій пошукової машини

- **Spider** («павук») - програма, яка завантажує в пошукову машину Web-сторінки. Працює аналогічно браузеру, встановленому на комп'ютері користувача, але нічого не відображає ні на якому екрані. Якщо Ви хочете мати уявлення про те, що саме завантажує в пошукову систему «павук», відкрийте якусь Web-сторінку і виберіть в меню Вид браузера пункт Перегляд HTML (або «вихідного») коду.

- **Crawler** («Черв'як», або «мандрівний павук») - програма, здатна знайти на Web-сторінці всі посилання на інші сторінки. Її завдання - визначити, куди далі повинен повзти «павук», керуючись посиланнями або заздалегідь заданим списком адрес.

Перші дві програми, що працюють «в зв'язці», часто називають пошуковий робот (а іноді - HTTP-робот).

Залежно від призначення розрізняють такі види пошукових роботів:

- національні, або головні. Збирають інформацію з одного національного домену, наприклад, .ua або .su, і прийнятих до індексації сайтів;

- глобальні. Збирають данні з усіх національних сайтів;
- індексатори картинок, аудіо та відео файлів;
- дзеркальники. Визначають дзеркала ресурсів;
- довідкові. Підраховують число посилань на сайті;
- підсвітчики. Оформляють результати пошукових систем, наприклад, виділяють в тексті запитувані словосполучення;
- перевіряючі. Контролюють наявність ресурсу в базі даних пошукової системи і число проіндексованих документів;
- стукачі (або дятли). Періодично визначають доступність сайту, сторінки або документу, на який веде посилання;
- шпигуни. Виконують пошук посилань на ресурси, ще не проіндексовані пошуковими системами;
- доглядачі. Запускаються в ручному режимі і перевіряють отримані результати;
- дослідники. Використовуються для налагодження пошукових алгоритмів і вивчення окремих сайтів;
- швидкі роботи. В автоматичному режимі перевіряють дату останнього оновлення і оперативно індексують нову інформацію.

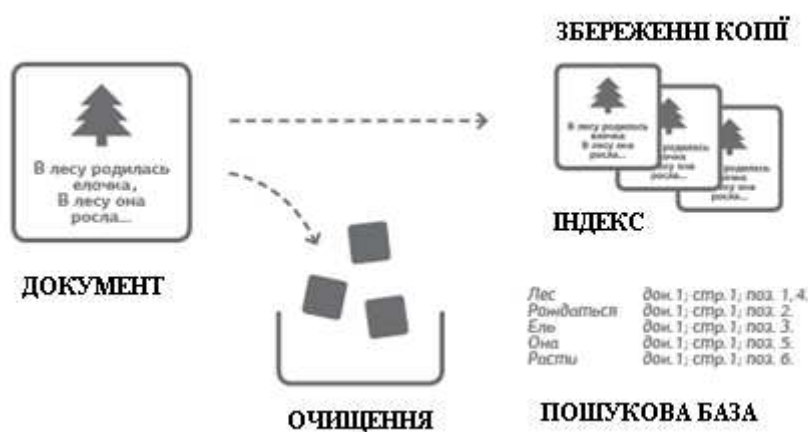


Рисунок 2.3 – Процес індексації

- **Indexer** (індексатор) - програма, яка «розбирає» сторінку на складові частини і аналізує їх. Виділяються і аналізуються заголовки Web-сторінок, заголовки документів, посилання, текст документів, окремо текст, виділений напівжирним шрифтом, курсивом і т.п.

Для зручності обробки даних пошукова система заносить всі знайдені в Інтернеті сторінки в індекс, який допомагає стиснути їх обсяги для оптимального зберігання. Щоб розуміти, за рахунок чого можливе скорочення обсягу інформації, необхідно знати, як документ представляється в пошуковій системі.

Спочатку відбувається очищення сторінки від різних нетекстових елементів, таких як графіка, HTML-теги і т.п. В результаті залишається «чистий» текст для подальшої обробки.

Далі всі слова з тексту розташовуються в алфавітному порядку, а всі елементи, які словами не є (пробіли, знаки пунктуації та інше), відкидаються. При цьому пошукова машина не заносить в індекс слова в тій формі, в якій вони наведені в тексті. За допомогою алгоритму лінгвістичної обробки всі слова представляються у початкових граматичних формах або основах. Це дозволяє скоротити місце в індексі і зробити пошук більш точним. З оброблених основ складається подоба словника, де вказується адреса сторінки і конкретне місце розташування кожної основи (номер входження). У пошуковому індексі зберігаються тільки номери основ, а самі основи розташовуються окремо. Якщо на сторінці знаходиться кілька входжень одного слова, то в індексі вказується номер цієї сторінки і все входження цього слова на ній. Виходить щось на зразок зворотної копії всіх сторінок Інтернету. Її називають інверсним або інвертованим індексом.

Пошукові машини зберігають і прямий індекс. Прямий індекс являє собою стислу текстову копію всіх сторінок Інтернету. Це значно економить час, наприклад, при показі цитат. Збережена копія сайту - це сторінки в прямому індексі пошукової системи.

- **Database** (база даних) - сховище всіх даних, які пошукова система завантажує й аналізує. Вимагає величезних ресурсів як для зберігання, так і для подальшої обробки.

- **Search Engine Results Engine** (система видачі результатів пошуку) вирішує, які сторінки задовольняють запиту користувача і в якій мірі. Саме з цією частиною пошукової системи «спілкується» користувач.

Щоб зрозуміти, що саме шукає користувач, пошукова система проводить ретельний лінгвістичний аналіз запиту. Спочатку визначається мова, на якій було сформульовано запит.

Далі проводиться робота по трактуванню морфології. Пошукова система розрізняє не тільки слова із запиту в усіх їх морфологічних формах, а й синоніми, проте при ранжируванні перевага віддається точному входженню.

Також пошуковим системам доводиться розмежовувати омоніми (слова з однаковим написанням, але різним значенням). Наприклад, одне і те ж слово може бути витлумачено і як дієслово, і як іменник.

Отже, пошукова машина, отримавши запит на пошук, не відправляється в тривалу подорож по «Всесвітній павутині», як вважають деякі користувачі, а аналізує лише ту інформацію, яку зібрала раніше. З одного боку, це дозволяє різко підвищити швидкість обробки запиту на пошук. З іншого, обмежує область пошуку внутрішніми ресурсами пошукової системи, які, по-перше, обмежені (жодна пошукова машина не в змозі завантажити в свою базу даних інформацію з усіх вузлів мережі), по-друге, вже в якійсь мірі застаріли. Ситуація в Інтернет змінюється дуже швидко. Якщо «павук» з метою оновлення інформації про вже проіндексовані один раз Web-сторінки «заповзає» на них раз у два місяці, користувач ризикує отримати в результатах запиту посилання на вже неіснуючу Web-сторінку.

Процес завантаження з Мережі інформації та попереднього аналізу її пошуковою машиною називається **індексація** (рис. 2.3), а сама база даних пошукової машини, в якій зберігається зібрана інформація - **індекс**.

Глибина індексації може бути різною. Повні тексти документів, розміщених на сторінці, в базу даних копіюють не всі пошукові роботи - деякі обмежуються лише заголовками. Коли користувач формує запит на пошук, пошукова машина переглядає свою базу даних і видає перелік Web-сторінок, що містять слова, введені користувачем в поле введення (їх часто називають ключові слова). Таких сторінок може бути дуже багато. Завдання пошукової машини - відібрати ті з них, які найбільшою мірою відповідають запиту користувача (тобто релевантні йому) і вказати посилання на них в числі перших.

Висока швидкість пошуку забезпечується не тільки за рахунок того, що пошукова машина звертається до вже зібраної і тієї, що зберігається тут же, у неї «під рукою», інформації. Аналізуючи зібрані дані, пошукова машина виконує індексацію бази даних, в процесі якої кожному слову ставляться у відповідність його «координати» - номер документа, в якому є дане слово, а найчастіше і позиція слова в документі (номер позиції і номер слова в ньому).

Алгоритмом пошуку можна назвати метод, керуючись яким пошукова машина приймає рішення, включати чи не включати посилання на сторінку або документ у результати пошуку.

Майже кожна пошукова машина використовує свій власний алгоритм пошуку, і його деталі являють собою ноу-хау розробників пошукача.

2.3 Ранжування інформації пошуковими системами.

Ранжування - це формування упорядкованого переліку пошукової видачі за результатами запитів.

Кожна пошукова система використовує свої алгоритми ранжування.

Алгоритми ранжування - принципи і послідовності оцінки сторінки Інтернет-ресурсу роботом пошукової системи для видачі в результатах пошуку під певний запит. Подібні алгоритми постійно створюються і оновлюються

програмістами пошукачів і визначають, який сайт вийде в пошуку вище, а який нижче, за якими ознаками.

Основні фактори, на які звертають увагу алгоритми ранжування більшості пошукових систем. [7]

Фактори домену (Домен являє собою якусь обмежену частину простору ієрархічних імен вузлів Всесвітньої мережі Інтернет, яка має централізоване адміністрування і обслуговується через сервери системи доменних імен (DNS, Domain Name System). Кожен домен ідентифікується через власне ім'я):

1. Вік домену. У деяких випадках більш старі домени можуть мати перевагу перед більш молодими;

2. Історія домену. "Старі гріхи" можуть надовго заплямувати репутацію домену;

3. Дані про власника домену. Закриті дані можуть бути підозрілим сигналом. Крім того, якщо пошукова система знає, що власник домену не чистий на руку, це може мати прямий вплив на ранжування;

4. Термін реєстрації. Якщо домен зареєстрований і сплачений на тривалий термін (більше року), це свідчить про серйозні наміри власників;

5. Ключові слова в домені. Позитивно впливають на релевантність;

6. Регіональність домену. Домени .ua матимуть перевагу для української видачі. Відповідно, ранжуватися гірше за межами України.

Фактори на рівні сторінки:

1. Ключові слова в контенті, щільність, порядок. Зрозуміло, що ключові слова в контенті критично важливі. Однак, з розвитком пошукових алгоритмів значення даних чинників змінюється: пошукові роботи вже далеко не прямолінійні, навчилися не тільки вважати ключі, але і розуміти сенс контенту. Зараз якість контенту набагато важливіша за кількість і щільності ключових слів. Порядок ключових слів важливий, тому що може мати прямий вплив на зміст запиту;

2. Довжина контенту. Довший контент має перевагу. Мається на увазі, що чим повніше розкрита тема, якою цікавиться користувач, тим краще;

3. Прихований семантичний показник (LSI). Крім основних ключових фраз, є інші слова, які допомагають пошуковій машині зрозуміти, що на сторінці мова йде саме про цю тему, а не про інший схожою. Наприклад, наявність слів "Україна", "Чорне море" дають зрозуміти, що на сторінці йдеться про українську Одесі, а не про американську;

4. Швидкість завантаження сторінки. Один з найбільш критично важливих чинників ранжирування на сьогоднішній день. У століття мобільних технологій сторінки повинні завантажуватися дуже швидко. Повільні ранжируватимуться гірше;

5. Медіа матеріали. Наявність фото і відео на сторінці - плюс. Зображення повинні бути оптимізовані (швидко завантажуватися, мати ключові слова в описі);

6. Частота оновлення контенту. Коли останній раз оновлювався контент на сторінці, наскільки часто це відбувається - ці відомості можуть впливати на релевантність за певними типами запитів;

7. Якість вихідних посилань. Позитивно впливають на релевантність посилання на авторитетні ресурси по темі сторінки. І навпаки, посилання на сайти, які тематично далекі і не є авторитетними, тягнуть сторінку вниз;

8. Кількість вихідних посилань. Як правило, чим більше посилань йде зі сторінки, тим гірше для неї;

9. Кількість і якість внутрішніх посилань. Чим більша кількість сторінок посилаються на дану сторінку і чим вони більш авторитетні в рамках сайту, тим більше ваги отримує дана сторінка;

10. Унікальність контенту. Оригінальний контент, безумовно, важливий і потрібний. Але слід пам'ятати про пріоритет корисності контенту. Унікальність - поняття відносне, головне завдання контенту - бути корисним для користувачів;

11. Корисні додаткові матеріали. Маються на увазі будь-які елементи сторінки, які створюють додаткову користь для відвідувачів. Це може бути онлайн-калькулятор вартості, конвертер величин, 3D-візуалізація тощо;

12. Форматування, нумеровані і маркіровані списки. Форматування служить для поліпшення сприйняття тексту користувачами, і пошукові машини це враховують. Тому заохочуються добре структуровані сторінки, які мають наявності нумеровані і маркіровані списки. Зручність сайту - поняття суб'єктивне. Як пошукові системи можуть знати, зручний сайт для користувачів чи ні? За допомогою збору та аналізу даних про призначеному для користувача взаємодії. У загальному випадку, якщо користувачі проводять багато часу на сайті, переглядають велику кількість сторінок, повертаються знову - це явно позитивні сигнали. А якщо переважна більшість відвідувачів закривають сайт, ледь зайшовши на нього, це вже сигнал негативний.

Фактори сайту:

1. Оптимізована сайту під мобільні пристрої. Крім того, що це додаткова зручність для мобільних користувачів, частка яких стрімко зростає, даний показник з квітня 2015 року є офіційно оголошеним фактором ранжирування;

2. Архітектура сайту. Пошукові роботи цінують логічність архітектури не менше (а, можливо, і більше), ніж люди;

3. Наявність карти сайту. Фактор, відомий настільки давно, що може вважатися класикою. Карти сайту (в них містяться відомості про структуру сайту і всіх вхідних в нього сторінок) допомагають пошуковим роботам краще "розуміти" сайт і швидше індексувати;

4. Репутація, відгуки користувачів. Пошуковики відстежують не тільки посилання на сайт, але і згадки про нього. Позитивний або негативний посил, яким супроводжується згадка, також може вплинути на ранжування. Як і в інших випадках, в першу чергу будуть враховуватися відгуки на трасових ресурсах.

Поведінкові чинники:

Поведінкові чинники ранжирування - набір показників взаємодії користувачів з сайтом, який використовується пошуковими системами як один із способів оцінки сайту і визначення його позиції у видачі.

1. Час на сайті - загальний час, який відвідувач проводить на сайті, до закриття останньої вкладки;
2. Глибина перегляду - кількість сторінок, які користувач переглядає протягом однієї сесії;
3. Час перегляду сторінки - час знаходження на сторінці, з моменту потрапляння на неї до моменту виходу;
4. Задоволені переходи - перехід по внутрішньої засыланні після 30 секунд перебування на сторінці;
5. Показник відмов - кількість (відсоток) відвідувачів, які не просунулися в перегляді сайту далі однієї сторінки;
6. Повернення відвідувачів - повторні відвідування одного і того ж користувача.

Спеціальні чинники

Існує ряд факторів, які не належать безпосередньо до оцінки сайтів, які ранжуються, а описують особливості роботи алгоритмів:

1. Персоналізована видача. "Бути в топі", "Позиція запиту" - поняття дуже відносні, якщо не сказати абстрактні, про що більшість людей навіть не підозрює. Google збирає, аналізує історію поведінки користувачів в інтернеті і виходячи з цього коригує пошукову видачу. Також має місце прив'язка до регіону, в якому знаходиться користувачі і сайти. Таким чином, сам факт постійного знаходження "в топі" або фіксації позицій неможливий - при кожному введенні пошукового запиту відбувається новий запуск алгоритму ранжирування, який враховує не тільки характеристики вашого сайту і сайтів конкурентів, але і персоналізацію під шукає користувача;

2. QDF (Query Deserves Freshness - запит заслуговує свіжості). Для певних типів запитів, для яких, на думку Google, більш релевантною буде свіжа інформація, перевага віддається новому контенту.

3. QDD (QDD - Query Deserves Diversity - (дослівно) запит заслуговує різноманітності): у випадках, коли не зовсім зрозуміло, який тип контенту

шукає користувач, Google додає різноманітності, щоб вгадати бажання користувача.

2.4 Найбільш популярні пошукові системи світу та їх особливості

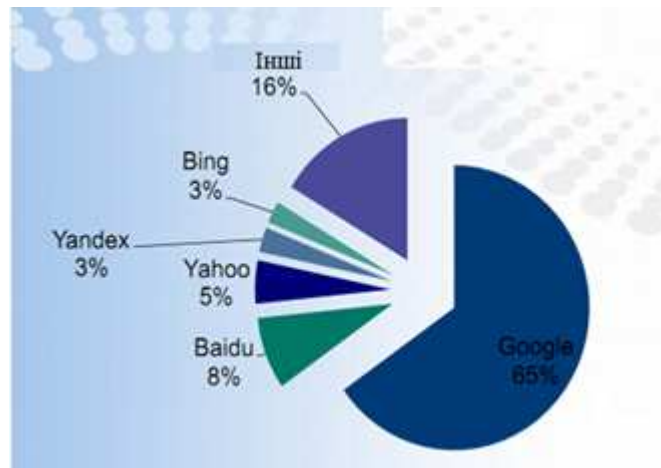


Рисунок 2.4 – Доля оброблюваних запитів (статистика за 2017 рік).

Перше місце зі значним відривом займає і міцно утримує Google, обробляючи до 70% від загальної кількості запитів, що становить близько 115 мільярдів запитів (рис. 2.5).

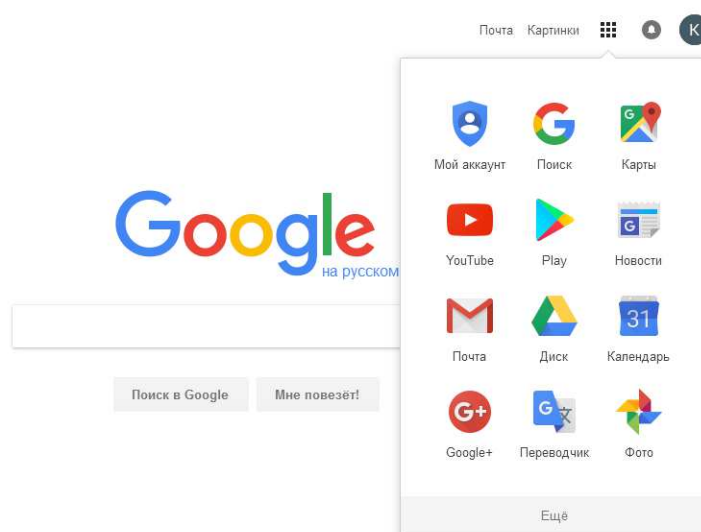


Рисунок 2.5 – Інтерфейс першої сторінки пошукової системи Google

Назва пошукової системи походить від зміненого слова googol (гугол), яке означає величину 10^{100} , тобто одиницю зі ста нулями.

Пошукова система, запущена в 1998 році і є нині одноосібним лідером серед глобальних пошукових систем по всім значимим параметрами. Головна перевага Google – обсяг його індексного файлу, який становить на сьогодні понад 4,2 мільярди web-сторінок і статей з груп новин за інтересами. В добу програми-роботи системи індексують близько трьох мільйонів нових та оновлених сторінок, при тому, що актуалізація бази проводиться кожні 28 днів.

Друга безсумнівна перевага Google - його здатність індексувати документи не тільки у вигляді HTML-файлів, але також документи в форматах PDF, RTF, PS, DOC, XLS, PPT, WP5 та ряду інших. При цьому Google дозволяє миттєво конвертувати сторінки в зазначених форматах в звичайний HTML-файл, що звільняє користувача від необхідності мати спеціальне програмне забезпечення для доступу до файлу.

Наступна перевага - спеціально розроблений модуль ранжирування результатів - PageRank. Він заснований на алгоритмі, згідно з яким спочатку встановлюється структура посилань у всій мережі, а потім кожна окрема сторінка ранжується відповідно до числа і значимості посилань на неї з інших сторінок. При цьому авторитетність зовнішніх посилань важливіша, ніж їх кількість. Подібний алгоритм дозволяє істотно підвищити релевантність посилань, внаслідок чого Google відрізняє високий ступінь відповідності знайденої інформації інтересам користувача. Цей результат досягається, зокрема, ще й за рахунок спеціальної підсистеми захисту користувача від сайтів, які просуваються за допомогою різних недобросовісних методів.

2-е місце – пошукова система Китаю Baidu, на частку якої припадає 8% від загального обсягу запитів, а саме - 14,5 мільярдів запитів (рис. 2.6).



Рисунок 2.6 – Інтерфейс першої сторінки пошукової системи Baidu

Заснована в 2000 році, засновники - Робін Лі і Ерік Сю, які отримали вищу освіту в США. \$ 1,2 млн. стартового капіталу залучили від американських венчурних компаній. Назва компанії взяли з поеми часів династії Сун, буквально воно означає «сто раз». Через рік отримали від венчурних компаній ще \$ 10 млн. інвестицій. У 2004 році Baidu стала лідируючою пошуковою системою в Китаї. Перша модель роботи сервісу передбачала надання іншим ресурсам платформи для пошуку. При цьому Baidu отримувала гроші за кожен запит. Проте, ця модель незабаром набридла Робіну Лі, і він вирішив перетворити сервіс в незалежний проект.

Особливості пошукової системи Baidu:

1. На даний момент Baidu контролює 80% китайського ринку, що становить близько 465 млн. користувачів. Величезну частину трафіку (близько 40%) в Baidu забезпечує сервіс з пошуку прямих посилань на скачування аудіотреків;

2. Високий дохід пов'язаний, в першу чергу, з оригінальним способом монетизації. Baidu зробила платною можливість попадання сайту в топ пошуку. Вартість залежить від тематики ресурсу і ключових фраз;

3. Створено сервіс «Гніздо фенікса», за допомогою якого проводиться перевірка рекламодавців, надається допомога у створенні власних оголошень, а також відстежується трафік.

На даний момент Baidu продовжує вести боротьбу на два фронти. За кількістю сервісів вони нітрохи не поступаються Google. Серед них варто виділити наявність соціальної мережі, електронної пошти, онлайн-карт, стрічки новин, Baidupedia - яка, будучи місцевим аналогом онлайн-енциклопедії, включає в себе понад 7 млн. статей, що більше, ніж в англійській та німецькій розділах «Вікіпедії», разом узятих. Також надзвичайно цікавий підхід до реалізації онлайн-карт, де кожен користувач може подивитися 3D-панораму китайських міст. Запуск стількох сервісів був зроблений для забезпечення світової інтеграції Baidu.

Але, на жаль, китайському пошуковикові поки не вдалося скласти конкуренцію Google в інших країнах. Незважаючи на правильно обрану стратегію, спрямовану на початкове завоювання азійського ринку, Baidu заважають чинники, за рахунок яких було досягнуто домінування в Китаї. Виявилося, що користувачам не дуже подобається обмеження контенту, посилання на скачування пісень заборонені в інших країнах, а переклади пошукача на інші мови виконані недостатньо грамотно.

3-є місце - за пошуковою системою Yahoo!, місячний обсяг у якій становить 8,6 мільярди оброблених запитів (рис. 2.7).

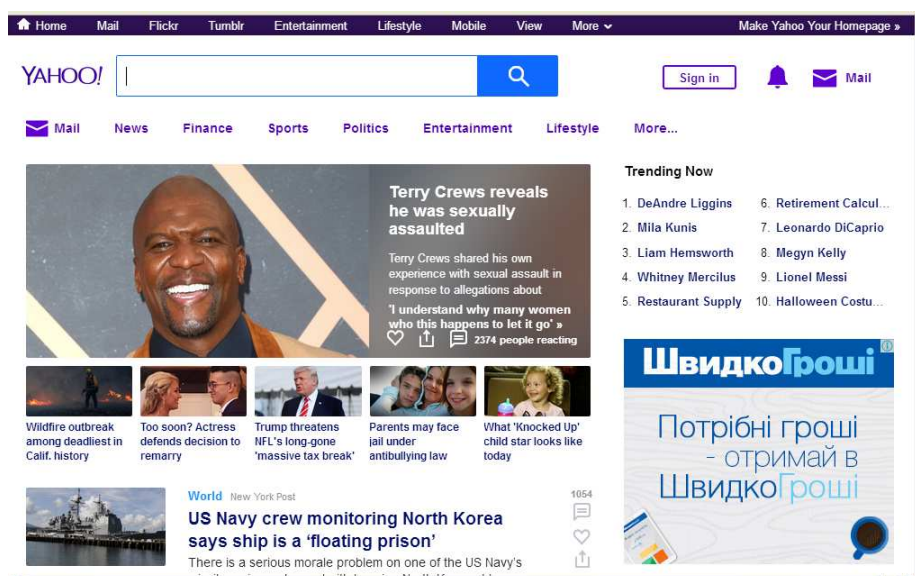


Рисунок 2.7 – Інтерфейс першої сторінки пошукової системи Yahoo!

У січні 1994 року аспіранти Стенфордського університету Девід Філо (англ. David Filo) і Джеррі Янг (англ. Jerry Yang) створили веб-сайт, який називався «Путівник Джеррі по Всесвітній Павутині». «Путівник» представляв собою каталог інших сайтів.

У квітні 1994 року сайт був перейменований в Yahoo!

Дані, які заносяться в каталог, проходять «ручну обробку». Сьогодні на Yahoo! працюють понад 200 редакторів, які щодня переглядають простір Інтернету в пошуках найбільш цінних ресурсів за темами, що викликають суспільний інтерес.

Каталог не засмічує свої рубрики посиланнями на сторінки-одноденки.

Користувачів завжди дратують посилання, що вказують на давно неіснуючі ресурси, тому Yahoo! скрупульозно підходить до формування своїх розділів.

Ретельність в підборі інформації забезпечує високу репутацію Yahoo!, незважаючи на те, що сукупний обсяг його ресурсів вкрай малий.

4-е місце – пошуковик від Microsoft - Bing, що обробляє близько 5 мільярдів запитів на місяць (рис. 2.8).

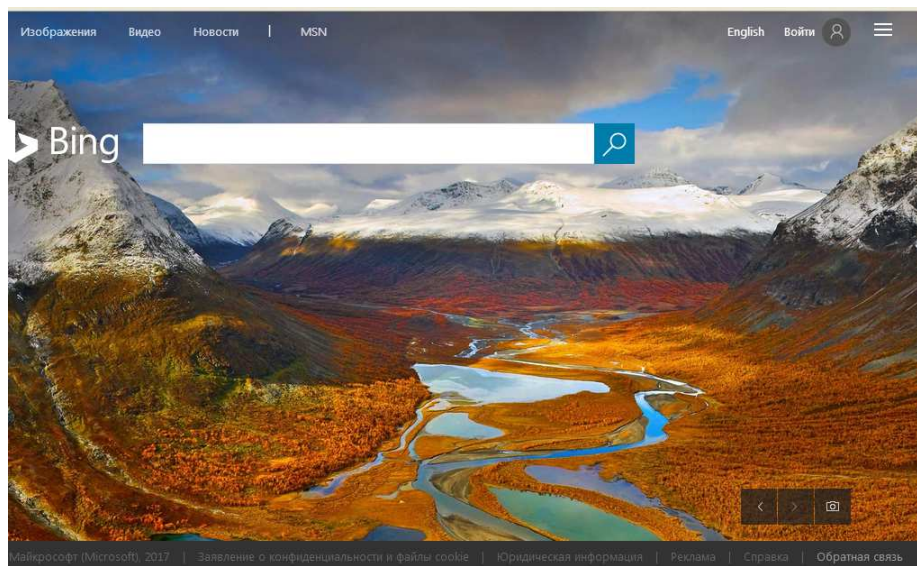


Рисунок 2.8 – Інтерфейс першої сторінки пошукової системи Bing

Зараз пошуковик Bing є четвертим по використанню в світі: велика заслуга в цьому - використання Internet Explorer за замовчуванням в Windows. Більшість користувачів просто поставили перед фактом - ось Explorer, а ось і Bing: користуйтеся на здоров'я.

Особливості:

1. У пошуковика Bing були три попередники, а точніше три інших назви: MSN Search, Windows Live Search, Live Search. І нарешті 1 червня 2009 року - з'явився він - пошуковик з ємною і звучною назвою – Bing;

2. Відразу після появи у Bing з'явилася жартівлива розшифровка - Bing Is Not Google;

3. В Bing щодня змінюється тема оформлення стартової сторінки;

4. Дуже зручний пошук відео - ви можете відбирати відео як по довжині, так і за якістю зображення, так і по джерелу. В американській версії пошуковика також є вбудований плеєр, який програв відео без переходу на джерело;

5. Також не поступається йому по зручності і пошук за картинками - там критеріїв ще більше;

6. Є можливість перегляду результатів пошуку на одній сторінці за рахунок нескінченного гортання. Це зручно;

7. Кодова назва пошукача Kumo. Все-таки не так сильно звучить як Bing;

8. Bing розвивається семимильними кроками і готовий потіснити трохи Google з пошукового олімпу.

5-е місце зайняв російський Yandex зі «скромним» показником 3 мільярди оброблюваних запитів (рис. 2.8).

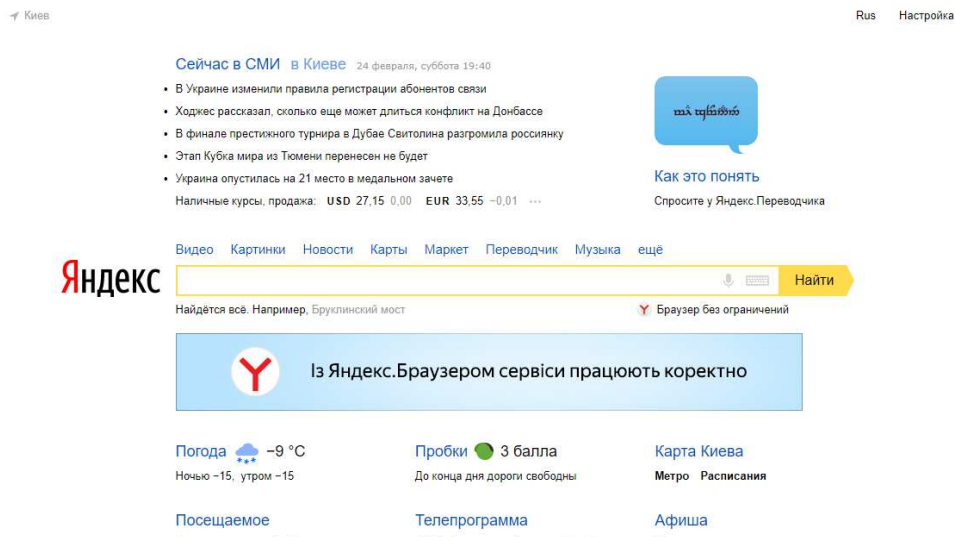


Рисунок 2.8 – Интерфейс першої сторінки пошукової системи Yandex

Яндекс був запущений 23 вересня 1997, але він не відразу став однією з найпопулярніших пошукових систем Рунета, хоча вигідно відрізнився від інших пошукачів. Його назва походить від англійського Yet another indexer, що можна дослівно перекласти як «черговий індексатор».

Яндексом проіндексовано понад 1.5 мільйонів російських і зарубіжних російськомовних серверів, а також серверів на території СНД (всього враховано близько 300 мільйонів оригінальних документів). Актуалізація бази здійснюється щотижня.

За рахунок вбудованої системи морфологічної обробки термінів Яндекс пристосований для формування запитів на природній російській мові. Найпотужніша лінгвістика дозволяє врахувати практично всі можливі відтінки вживання ключових слів і скласти пошуковий припис у вищій ступені широко, охопивши всі можливі поєднання термінів.

Безумовною перевагою Яндекс є модуль ранжирування результатів пошуку. В його основі лежить алгоритм аналізу місця розташування та повторюваності терміна в документі і аналіз зовнішніх посилань на сторінку. Система виробляє угруповання знайдених сторінок по сайтам на етапі формування списку результатів і видає по одному документу з сервера, який, на її думку, найбільш відповідає запиту. При цьому на запиті вказується

загальна кількість документів з даного сервера, які також містять шуканий термін. Яндекс показує знайдені терміни в навколишньому словниковому контексті, що дозволяє відразу ж встановити ступінь відповідності знайденого документа інформаційній потребі користувача.

Українські пошукові системи.

Український портал (<http://www.uaportal.com>).

Український портал відкрився 3 квітня 2000 року та був задуманий як центральна сторінка України, і покликаний служити точкою старту в Інтернет подорожах українського користувача. Портал побудований за принципом, що дозволяє навіть необізнаній людині шукати і знаходити інформацію в Інтернет.

На сторінках portalу публікуються матеріали, що розповідають про Інтернет і про його сервіси. Також відбираються ті кращі новини і сайти українського сегменту мережі, про які на думку керівників сайту буде цікаво дізнатися користувачам. Для цього відстежуються зміни на провідних сайтах України і розповідається про них в короткій (реферативній формі), з посиланнями на джерело.

Meta (<http://www.meta.ua/>).

META — український пошуковий портал в Інтернеті. По суті META.ua представляє широкопрофільний портал із зручною пошуковою системою. Пошукова система META дозволяє шукати по всьому українському Інтернету, а також по Реєстру українських сайтів з урахуванням української, російської та англійської морфології. Це означає, що незалежно від граматичної форми ключових слів, отримуються документи, які містять шукані слова у всіх формах.

Bigmir)net (<http://www.bigmir.net/>)

bigmir)net — російськомовний інтернет-портал заснований в Україні. Почав свою роботу в листопаді 2000 року. На порталі працюють 17 сервісів bigmir)net, найбільшими з яких є Пошук, Рейтинг сайтів, Пошта та Погода. Щодня портал відвідують понад півмільйона осіб.

2.5 Мета-пошукові системи

Багато хто вважає, що Google дає такі точні результати, що не потрібна ніяка інша пошукова система. Але це не зовсім так.

Ні в одній з пошукових систем немає індексу ВСІХ існуючих веб-сторінок. Насправді, ніхто не знає, наскільки простягається всесвітня павутина. У будь-якому випадку, абсолютно точно те, що основні пошукові системи індексують тільки частину "публічно індексуємої павутини". Крім того, кожна пошукова система індексує різні веб-сторінки, а це значить, що, користуючись тільки однією пошуковою системою, Ви пропустите відповідні запиту результати, які можна знайти в інших пошукових системах.

Єдиний спосіб зробити пошук по WWW більш ефективним - використовувати мета-пошукову систему.

Мета-пошукова система (також звана метакраулером або мульти-поточною системою) - це пошуковий інструмент, який посилає ваш запит одночасно на декілька пошукових систем, каталогів і, іноді, в так звану невидиму (приховану) павутину - збори онлайнової інформації, що не проіндексована традиційними пошуковими системами. Зібравши результати, мета-пошукова система (МПС) видаляє дубльовані посилання і, відповідно до свого алгоритму, об'єднує/ранжує результати в загальному списку.

На відміну від окремих пошукових систем і директорій, мета-пошукові системи:

- не мають власних баз даних;
- не реєструють URLи сайтів (універсальний локатор ресурса-застосовується для позначення адрес майже всіх ресурсів Інтернету)

Переваги: МПС економлять досить багато часу, позбавляючи від необхідності вводити запит в кожній окремій пошуковій системі. Результати, в більшості випадків - у вищій мірі релевантні. МПС можуть використовуватися веб-майстрами для з'ясування, чи присутній їх сайт в головних пошукових системах, які у нього рейтинги і наскільки популярні посилання.

Недоліки: Так як деякі пошукові системи і каталоги не підтримують додаткові технології пошуку, такі як лапки, в які полягають фрази, або Булеві оператори, при застосуванні таких технологій в результатах МПС не будуть присутні результати з таких пошукових систем, або ці результати будуть невідповідними.

2.6 Основи роботи з пошуковою системою Google.

2.6.1 Особливості пошуку в Google.

Оскільки Google на сьогоднішній день є безумовним лідером серед пошукових систем, слід розглянути особливості роботи з даним пошуковиком та основні оператори, які дозволяють покращити пошук.

Найпростіший спосіб пошуку в Google – надрукувати один чи більше пошукових термінів (слів або фраз, які найкращим чином описують інформацію, що вам потрібна) у вікні для пошуку і натиснути 'Enter'. У відповідь Google поверне сторінку результатів пошуку: перелік веб-сторінок, які відповідають Вашим пошуковим термінам. При цьому сторінки, які відповідають пошуковим термінам найкраще, виводяться на початку цього переліку. Зрозуміло, що вірний вибір пошукових термінів – це ключ до відшукування потрібної інформації.

Пошукова система Google НЕ є регістро-залежною. Усі літери, як би Ви їх не вводили, автоматично переводяться до нижнього регістру (стають маленькими). Тому, наприклад, результати пошуку для запитів Державний Університет Телекомунікацій, державний університет телекомунікацій та ДержАвний УНІверситет тЕЛЕкомуНІКАцій будуть однаковими.

За замовчуванням, Google повертає лише ті сторінки, які включають усі Ваші пошукові терміни. Тому немає потреби вставляти "and" між термінами. Але візьміть до уваги, що порядок слів у запиті може впливати на результати пошуку.

Також, не слід перейматися через правопис. Засіб перевірки правопису Google автоматично використовує найбільш уживану форму певного слова, незалежно від того, чи Ви його правильно написали.

Зміна форми слова.

Google використовує технологію зміни форми слова. Тому, коли це може бути доречним, Він шукає не лише пошукові терміни, задані Вами, але також і слова, схожі на усі або деякі з Ваших пошукових термінів. Якщо Ви задаєте параметри пошуку як потреби домашніх тварин, Google також шукатиме результати для потреби у домашніх тваринах та інших можливих варіантів Ваших пошукових термінів. Будь-які варіанти Ваших пошукових термінів, по яких здійснювався пошук, виділяються у фрагменті тексту, що супроводжує кожний результат пошуку.

"Мені пощастить" ("I'm Feeling Lucky").

Після того, як Ви ввели у форму пошуку Ваші пошукові терміни, Ви можете спробувати натиснути кнопку "Мені пощастить", яка одразу перенаправляє браузер на веб-сторінку, що найкраще (за версією Google) відповідає Вашим пошуковим термінам. Тобто у цьому разі Ви оминете сторінку результатів пошуку. Якщо ж Ви скористаєтеся звичайною кнопкою "Пошук Google", тоді посилання, яке відповідає опції "Мені пощастить", буде першим у переліку результатів пошуку.

Наприклад, якщо Ви шукаєте сайт Державного університету телекомунікацій, Ви можете задати у формі для пошуку ДУТ і натиснути "Мені пощастить" замість кнопки "Пошук Google".

Google автоматично перенаправить Ваш браузер на "<http://www.dut.edu.ua>"

Налаштування пошуку.

Змінити налаштування пошуку за замовчуванням Ви можете, якщо перейдете по посиланню Налаштування (Preferences) (<http://www.google.com/preferences?hl=en> для англomовної версії, <http://www.google.com.ua/preferences?hl=uk> для української).

Існує можливість задавати мову сторінок, які включаються в результати пошуку. Такий спосіб пошуку не можна вважати універсальним, оскільки він охоплює лише частину мережі, тому Google зазвичай пропонує за замовчуванням пошук усіма мовами. Проте вибір конкретної мови може бути корисним, якщо це принципова умова для Вашого пошуку. Зверніть увагу, що за замовчуванням, якщо Ви користуєтеся українською версією Google, Ви легко можете обрати результати лише українською мовою (ця опція пропонується на сторінці з результатами пошуку), але через загальну сторінку налаштувань ви можете обрати будь-яку мову з представлених там. Так само на сторінці налаштувань Ви можете обрати мову інтерфейсу для Google.

За замовчуванням, результати пошуку виводяться на сторінку по 10. Це значення гарантує найшвидшу відповідь. Але за Вашим бажанням Ви можете змінити це значення на 20, 30, 50 або 100.

Коли Ви шукаєте інформацію за допомогою Google і переходите по одному з посилань-результатів пошуку, нова сторінка заміщує собою сторінку з результатами пошуку. Якщо Вам зручніше, щоб результати пошуку залишались у відкритому вікні, оберіть опцію «Показати результати пошуку в новому вікні».

Складний пошук.

Складний пошук дає можливість налаштувати певну кількість опцій для того, щоб Ваш пошук був більш точним, а результати – більш корисними. Ви можете потрапити на цю сторінку по посиланню.

Складний пошук (Advanced Search) з домашньої сторінки Google.

Складний пошук дозволяє набагато більше, ніж просто перелічувати пошукові терміни. Ви можете шукати лише сторінки:

- які утримують УСІ ваші пошукові терміни;
- які утримують точну фразу, введену вами;
- які утримують, як мінімум, одне зі слів, введених вами;
- які НЕ утримують жодного зі слів, введених вами;
- створені заданою мовою;

- створені у заданому форматі файлу;
- ті, які оновлювалися протягом заданого проміжку часу;
- які утримують числа всередині заданого проміжку;
- розташовані у рамках заданого домена або веб-сайта;
- які не утримують матеріалів "для дорослих".

Мітка "Перевірка фактів" у результатах пошуку.

Іноді під час пошуку в Google з'являються результати з міткою "Перевірка фактів". У цих результатах Ви можете прочитати популярні твердження, пов'язані з Вашим пошуковим запитом, і дізнатися, правдиві вони, неправдиві чи частково правдиві (за результатами перевірки фактів видавцем).

Якщо видавець сайту перевірів достовірність певного твердження, пов'язаного з Вашим пошуковим запитом, відповідний результат пошуку міститиме таку інформацію:

- твердження, достовірність якого перевірялася;
- автор твердження;
- ім'я видавця, який перевіряв факти;
- підсумок перевірки фактів.

У Новинах Google

Статті, у яких перевірено достовірність тверджень, мають позначку "Перевірка фактів".

2.6.2 Найбільш вживані оператори пошуку.

Ви можете поліпшити результати пошуку, додаючи "оператори" до Ваших пошукових термінів у звичайній формі для пошуку Google або обираючи їх на сторінці складного пошуку. Google підтримує декілька операторів складного пошуку (спеціальних операторів), які по суті є словами запиту, що мають визначене значення для Google. Як правило, ці оператори змінюють правила пошуку певним чином або визначають зовсім інший тип пошуку. Більшість зі спеціальних операторів доступна зі сторінки складного пошуку, але не усі.

Деякі з більш уживаних операторів задаються не словами, а знаками пунктуації або не потребують двокрапки після себе. Серед цих операторів: OR (оператор АБО), "" (оператор лапок), - (оператор мінус) та + (оператор плюс).

Іноді Вам будуть потрібні лише результати, які включають деяку фразу (або словосполучення) повністю. В такому разі необхідно просто узяти Ваші пошукові терміни в лапки. Пошук по фразах є найбільш ефективним, якщо Ви шукаєте власні імена ("Тарас Шевченко"), фрази з віршів або пісень ("Чого являєшся мені у сні"), або інші відомі висловлювання ("з усіх втрат втрата часу найтяжча").

Пошук "+"

Google ігнорує дуже часто вживані слова і символи, такі як "where" (англ. "де"), "how" (англ. "як"), "the" (артикль), так само як деякі окремі цифри і окремі літери, тому що, як правило, вони уповільнюють пошук і при цьому не покращують його результати. Якщо якийсь з Ваших пошукових термінів проігноровано, Ви можете дізнатися про це одразу на сторінці результатів пошуку.

Але трапляється, що таке дуже часто вживане слово є важливим для отримання потрібних Вам результатів. В такому разі Ви можете включити його у запит, додавши знак "+" перед цим словом.

(Не забудьте залишити пробіл перед знаком "+"). Інший спосіб полягає у використанні пошуку по фразам. При пошуку по фразам загальновживані слова (наприклад, "where are you" (англ. "де ти є")) не виключаються.

Наприклад, для пошуку Star Wars, Episode I, можна використати Star Wars Episode +I АБО "Star Wars Episode I".

Виключення пошукових термінів. Оператор «-»

Якщо Ваш пошуковий термін має більше ніж одне значення (наприклад, океан може стосуватися географічних об'єктів, великого обсягу чогось або бути частиною назви популярної групи "Океан Ельзи"), Ви можете сфокусувати Ваш пошук, поставивши знак "-" перед словами, пов'язаними з тим значенням, яке Вас не цікавить. Наприклад, для того, щоб знайти інформацію про океан як такий, а не про музичну групу, можна використати такі параметри пошуку: океан –ельзи.

Примітка: якщо Ви включаєте у пошук негативний пошуковий термін, не забувайте залишати пробіл перед знаком "-".

Пошук синонімів.

Для того, щоб пошук відбувався не лише по заданих Вами пошукових термінах, але й по їхніх синонімах, треба поставити знак тільди ("~") безпосередньо перед відповідним пошуковим терміном.

Наприклад, знайти інформацію про факти, пов'язані з їжею та інформацією про приготування (англійською мовою) можна так: ~food ~facts (Нажаль, для української мови ця опція поки що не діє.)

Пошук "OR" ("Або")

Щоб знайти сторінки, які містять будь-який з двох пошукових термінів, у запиті треба вставити між цими термінами OR (великими літерами!) Наприклад, шукати інформацію про відпочинок у Лондоні чи Парижі можна так: відпочинок Лондон OR Париж

Пошук чисел всередині заданого проміжку.

Треба знайти щось, пов'язане з числами? Використайте у якості пошукового терміну два числа, розділених двома крапками без пробілів. Ви можете скористатися цим для пошуку довільних значень пов'язаних з числами: від дат (Willie Mays 1950..1960) до вимірів (500..1000 мм довжина). Але не забудьте вказати одиницю вимірювання або якийсь інший індикатор того, що саме являють собою ці числа. Наприклад, пошук DVD-плеєра ціною від \$50 до \$100 можна задати так: DVD player \$50..\$100.

@ и # – символи для пошуку по тегам і хештегам соціальних мереж.

Таблиця 2.1 – Інформаційно-пошукові системи. Підтримувані оператори.

Назва	AND (логічне И)	OR (логічне ЧИ)	NOT (логічне НЕМАЄ)	NEAR(оператор контекстної близкості)	Фраза	Пропуск слова	Доп. Параметри
Google	пробіл	OR	знак мінус	–	лапки	Символ «*»	+
Alta Viata	AND/&	OR/	NOT/!	NEAR/~	лапки	Символ «*»	+
HotBot	AND/&/+	OR/	NOT/~	NEAR	лапки	Символ «*»	+
Lycos	AND/+	OR	знак мінус	–	лапки	–	+
Yahoo!	пробіл/+	–	знак мінус	–	лапки	Символ «*»	+
Excite	AND/+	OR	знак мінус	–	лапки	–	+
Yandex	&/пробіл/+		~	+	лапки	Символ «*»	+
Апорт	I/&/+пробіл	ЧИ	НЕМАЄ	+	лапки	Символ «*»	–
Rambler	AND/&	OR/	NOT/!	+ (число, запит)	лапки	–	+

2.6.3 Найбільш популярні спеціальні оператори Google

cache: Запит [cache:] виведе версію веб-сторінки, яка зберігається у кеші Google. (Кеш пошукової системи – це записана під час індексації версія веб-сторінки. Оскільки сторінка може змінюватися після індексації, то текст, збережений у кеші, іноді відрізняється від тексту на реальній сторінці.) Наприклад, cache:www.google.com поверне кеш Google для домашньої сторінки Google. Зауважте, що між оператором "cache:" та адресою веб-сторінки не повинно бути пробілу!

Якщо Ви включите також інші слова у пошук, Google підсвітить ці слова у кешованому документі. Наприклад, `cache:www.google.com web` покаже кешований контент з підсвіченим словом "web". Аналогічний результат можна отримати, клікнувши по посиланню "Cached" ("Кеш") на основній сторінці результатів пошуку Google.

link: Запит `[link:]` виведе перелік веб-сторінок, які утримують посилання на зазначену веб-сторінку. Наприклад, `link:www.google.com` виведе перелік веб-сторінок, які утримують посилання на домашню сторінку Google. Зауважте, що між оператором "link:" та адресою веб-сторінки не повинно бути пробілу! Аналогічний результат можна отримати зі сторінки складного пошуку через Page Specific Search > Links (Пошук відомостей про сторінку > Посилання).

related: Запит `[related:]` виведе перелік веб-сторінок, "подібних" до зазначеної веб-сторінки. Наприклад, `related:www.google.com` виведе перелік веб-сторінок, схожих на домашню сторінку Google (тобто домашні сторінки інших пошукових систем). Аналогічний результат можна отримати зі сторінки складного пошуку через Page Specific Search > Similar (Пошук відомостей про сторінку > Подібні).

info: Запит `[info:]` видасть у відповідь інформацію, яку Google має стосовно зазначеної веб-сторінки. Наприклад, `info:www.google.com` видасть інформацію про домашню сторінку Google. Аналогічний результат можна отримати, якщо у явному вигляді надрукувати адресу веб-сторінки у пошуковій формі Google.

define: Запит `[define:]` поверне визначення слова, введеного після оператора, узяті з різних онлайн-джерел. Означення буде виведене для усієї фрази, введеної вами (тобто буде включати усі слова саме в тому порядку, як ви їх ввели).

site: Якщо Ви включите у свій запит оператор `[site:]`, Google обмежить можливі результати пошуку сайтами у означеному домені. Наприклад, `help site:www.google.com` буде шукати сторінки про допомогу (help) всередині `www.google.com`. А `help site:com` буде шукати сторінки про допомогу (help) на

сайтах, розташованих в домені .com. Аналогічний результат можна отримати зі сторінки складного пошуку через Advanced Web Search > Domains (Складний пошук > Домен).

allintitle: виводяться сторінки, у яких у заголовку тегу присутні вказані у запиті слова.

filetype – дозволяє знаходити файли тільки певного формату. *[Роль інформації у розвитку суспільства filetype:doc]*

intitle - Дозволить Вам відшукати тільки ті сторінки, в яких слово, написане після оператора, буде знаходитися в заголовку, а решта - в будь-якій частині сторінки. *[Фільми про тварин intitle: Дивитися безкоштовно]*

2.6.4. Як шукати за зображенням.

Зображення можна використовувати як запити, щоб шукати схожий вміст в Інтернеті.

Як працює пошук за зображенням.

Коли ви шукаєте за зображенням, результати можуть включати:

- подібні зображення;
- сайти, які містять це зображення;
- зображення, яке Ви шукаєте, в іншому розмірі.

Такий пошук дає найкращі результати, коли зображення популярне в Інтернеті. Тобто Ви отримаєте більше результатів для фотографій відомих місць, ніж для особистих зображень, як-от останніх фотографій Вашої сім'ї.

Під час пошуку зображень, відео або тексту в Google можна відфільтрувати матеріали, захищені авторським правом. Якщо Ви плануєте використовувати, публікувати або змінювати вміст, установіть фільтр "Права на використання" в Розширеному пошуку.

Як знайти зображення, текст або відео, які можна використовувати?

Перейдіть у Розширений пошук. Щоб знайти зображення, скористайтеся Розширеним пошуком зображень.

Введіть запит у полі "усі ці слова".

У спадному меню в розділі "Права на використання" виберіть потрібний тип ліцензії на вміст.

Типи прав на використання:

- вільне використання та розповсюдження. Можливість копіювати або розповсюджувати вміст у незміненому вигляді;

- вільне використання, розповсюдження та змінення. Можливість копіювати, змінювати або розповсюджувати вміст указаними в ліцензії способами;

- з комерційною метою. Якщо Ви шукаєте вміст для комерційного використання, виберіть цей параметр.

3 ОБРОБКА ІНФОРМАЦІЇ

Обробка інформації є основною функцією комп'ютера. Для того щоб числова, текстова, графічна і звукова інформація могла оброблятися на комп'ютері, вона повинні бути представлені у формі даних. При проектуванні технологічних процесів орієнтуються на режими їх реалізації. Режим реалізації технології залежить від об'ємно-часових особливостей вирішуваних завдань: періодичності і терміновості, вимог до швидкості обробки повідомлень, а також від режимних можливостей технічних засобів, і в першу чергу ЕОМ. Існують декілька видів обробки інформації такі як пакетний режим; режим реального масштабу часу; режим поділу часу; регламентний режим; запитальний режим; діалоговий режим; режим телеобробки; інтерактивний режим; однопрограмний режим; багатопрограмний режим (мультиобробка).

Обробка інформації полягає в отриманні одних "інформаційних об'єктів" з інших "інформаційних об'єктів" шляхом виконання деяких алгоритмів і є однією з основних операцій, здійснюваних над інформацією, і головним засобом збільшення її обсягу і різноманітності.

На самому верхньому рівні можна виділити числову і нечислову обробку. У зазначені види обробки вкладається різне трактування змісту поняття "дані". При числовій обробці використовуються такі об'єкти, як змінні, вектори, матриці, багатовимірні масиви, константи і т.д. При нечисловій обробки об'єктами можуть бути файли, записи, поля, ієрархії, мережі, відносини і т.д. Інша відмінність полягає в тому, що при числовій обробці зміст даних не має великого значення, в той час як при нечисловій обробки нас цікавлять безпосередні відомості про об'єкти, а не їх сукупність в цілому.

З точки зору реалізації на основі сучасних досягнень обчислювальної техніки виділяють наступні види обробки інформації:

- послідовна обробка, застосовувана в традиційній Фоннейманівській архітектурі ЕОМ, котра володіє одним процесором;

- паралельна обробка, застосовувана при наявності декількох процесорів в ЕОМ;
- конвеєрна обробка, пов'язана з використанням в архітектурі ЕОМ одних і тих же ресурсів для вирішення різних завдань, причому якщо ці завдання тотожні, то це послідовний конвеєр, якщо завдання однакові - векторний конвеєр.

3.1 Кодування інформації

Одним з процесів обробки інформації є її кодування. Для автоматизації роботи з різними даними (числовими, текстовими, звуковими та ін.) використовується прийом кодування - вираз одного типу даних через дані іншого типу. Так, природний людський мова являє собою систему кодування понять для вираження думок за допомогою мови, система кодування Брайля використовується у сліпих. Різні системи кодування успішно застосовуються в різних галузях техніки, науки, економіки.

Сигнали в комп'ютері передаються за допомогою електричних імпульсів. Щоб розрізнити кожну використовувану цифру десяткової системи, знадобилося б десять різних сигналів. З технічної точки зору чим менше видів сигналів, тим краще. Тому для електронного обчислювального пристрою ефективніше і зручніше двійкова система кодування - представлення чисел за основою 2, при якому значення виражаються комбінаціями 0 і 1. Простота здійснюваних операцій і можливість здійснювати автоматичну обробку інформації, реалізуючи лише два знаки, дають переваги, істотно перевищують недолік у вигляді швидкого зростання числа розрядів.

Кодування чисел. Числовою інформацію комп'ютер обробляє в двійковій системі числення. Таким чином, числа в комп'ютері представлені послідовністю цифр 0 і 1, званих бітами (біт - один розряд двійкового числа). На початку 1980-х рр. процесори для персональних комп'ютерів були 8-розрядними, і за один такт роботи процесора комп'ютер міг обробити 8 біт,

тобто максимально обробляється десяткове число не могло перевищувати 111111112 (або 25510). Послідовність з восьми біт називають байтом, тобто 1 байт = 8 біт. Потім розрядність процесорів росла, з'явилися 16-, 32- і, нарешті, 64-розрядні процесори для персональних комп'ютерів, відповідно зросла і величина максимального числа, оброблюваного за один такт.

Використання двійкової системи для кодування цілих і дійсних чисел дозволяє за допомогою 8 розрядів кодувати цілі числа від 0 до 255, 16 біт дає можливість закодувати понад 65 тис. Значень.

У ЕОМ застосовуються дві форми подання чисел:

- природна форма, або форма з фіксованою комою. У цій формі числа зображуються у вигляді послідовності цифр з постійним для всіх чисел становищем коми, що відокремлює цілу частину від дробової, наприклад +00456,78800; +00000,00786; -0786,34287. Ця форма незручна для обчислень і застосовується тільки як допоміжна для цілих чисел;

- нормальна форма, або форма з плаваючою точкою. У цій формі число виражається за допомогою мантиси і порядку як $N = \pm M \cdot P \pm r$, де M - мантиса числа ($|M| < 1$), r - порядок числа (ціле число), P - основа системи числення. Наведені вище цифри в нормальній формі будуть представлені як $+0,456788 \cdot 103$, $+0,786 \cdot 102$, $-0,3078634287 \cdot 105$.

Нормальна форма подання забезпечує великий діапазон відображення чисел і є основною в сучасних ЕОМ. Всі числа з плаваючою комою зберігаються в ЕОМ в нормалізованому вигляді. Нормалізованим називають таке число, старший розряд мантиси якого більше нуля.

У пам'яті ЕОМ для зберігання чисел передбачені формати: слово - довжиною 4 байта, півслова - 2 байта, подвійне слово - 8 байт.

Розрядна сітка для чисел з плаваючою комою має наступну структуру:

- нульовий розряд - це знак числа;
- з 1-го по 7-й розряд - записується порядок в двійковому коді;
- з 8-го по 31-й - вказується мантиса.

Кодування текстових даних. Двійкова система дозволяє кодувати і текстову інформацію. Вісім двійкових розрядів достатньо для кодування 256 різних символів.

Першим міжнародним кодом став стандартний 7-бітний код ASCII (American Standard Code for Information Interchange - американський стандартний код для обміну інформацією) (табл. 3.1). Поява даного коду в 1963 р зіграло значну роль, оскільки до цього різні комп'ютери просто не могли взаємодіяти один з одним. Кожен виробник по-своєму уявляв символи алфавіту, цифри і керуючі коди. В одних тільки апаратних засобах корпорації IBM використовувалося дев'ять різних наборів кодування символів. Але взаємодія між комп'ютерами стало нагальною потребою. У 1961 р комітет Американського національного інституту стандартів (ANSI), в якому була представлена велика частина виробників комп'ютерів, приступив до розробки міжнародного стандарту. Код ASCII став спільним знаменником для комп'ютерів, які раніше не мали один з одним нічого спільного. Всім буквам, цифрам, знакам і іншим символам (керуючим кодами) були поставлені у відповідність стандартні числові значення. Код ASCII підтримував 128 символів, що включають великі і малі символи латиниці, цифри, спеціальні знаки і керуючі коди. Базова таблиця кодування цього коду, починаючи з 32-го коду, наведена в табл. 1.2. Коди 0 ÷ 31 використані в даній таблиці як службові і керуючі.

Потім 7-бітний код ASCII був розширений до 256 символів і прийнятий як 8-бітний міжнародний стандарт ASCII-2, причому коди з 128 по 256 цього стандарту були задіяні для національних мов різних країн. Для СРСР в цій області була введена національна кодування ЯКІ-8 (код обміну інформацією, восьмизначний). Код ASCII залишився однією з небагатьох технологій, якій вдалося успішно пройти крізь десятиліття і дожити до наших днів. Сьогодні на основі коду ASCII випускається обладнання вартістю в мільярди доларів, більшість операційних систем до сих пір є сумісним з ASCII. Останнім стандартом в області кодування текстової інформації вважається 16-розрядний

універсальний міжнародний код Unicode (UNiversal CODE), що дозволяє кодувати 65 536 різних символів. Unicode охоплює 28 тисяч букв, знаків, складів і ієрогліфів національних мов світу, і 30 тисяч місць у ньому зарезервовано.

Таблиця 3.1 - Таблиця кодування ASCII

Бінарне	10-не	16-не	Гра-фічне	Бінарне	10-не	16-не	Гра-фічне	Бінарне	10-не	16-не	Гра-фічне
0010 0000	32	20	пробіл (_{SP})	0100 0000	64	40	@	0110 0000	96	60	`
0010 0001	33	21	!	0100 0001	65	41	A	0110 0001	97	61	a
0010 0010	34	22	"	0100 0010	66	42	B	0110 0010	98	62	b
0010 0011	35	23	#	0100 0011	67	43	C	0110 0011	99	63	c
0010 0100	36	24	\$	0100 0100	68	44	D	0110 0100	100	64	d
0010 0101	37	25	%	0100 0101	69	45	E	0110 0101	101	65	e
0010 0110	38	26	&	0100 0110	70	46	F	0110 0110	102	66	f
0010 0111	39	27	'	0100 0111	71	47	G	0110 0111	103	67	g
0010 1000	40	28	(0100 1000	72	48	H	0110 1000	104	68	h
0010 1001	41	29)	0100 1001	73	49	I	0110 1001	105	69	i
0010 1010	42	2A	*	0100 1010	74	4A	J	0110 1010	106	6A	j
0010 1011	43	2B	+	0100 1011	75	4B	K	0110 1011	107	6B	k
0010 1100	44	2C	,	0100 1100	76	4C	L	0110 1100	108	6C	l
0010 1101	45	2D	-	0100 1101	77	4D	M	0110 1101	109	6D	m
0010 1110	46	2E	.	0100 1110	78	4E	N	0110 1110	110	6E	n
0010 1111	47	2F	/	0100 1111	79	4F	O	0110 1111	111	6F	o
0011 0000	48	30	0	0101 0000	80	50	P	0111 0000	112	70	p
0011 0001	49	31	1	0101 0001	81	51	Q	0111 0001	113	71	q
0011 0010	50	32	2	0101 0010	82	52	R	0111 0010	114	72	r
0011 0011	51	33	3	0101 0011	83	53	S	0111 0011	115	73	s
0011 0100	52	34	4	0101 0100	84	54	T	0111 0100	116	74	t
0011 0101	53	35	5	0101 0101	85	55	U	0111 0101	117	75	u
0011 0110	54	36	6	0101 0110	86	56	V	0111 0110	118	76	v
0011 0111	55	37	7	0101 0111	87	57	W	0111 0111	119	77	w
0011 1000	56	38	8	0101 1000	88	58	X	0111 1000	120	78	x
0011 1001	57	39	9	0101 1001	89	59	Y	0111 1001	121	79	y
0011 1010	58	3A	:	0101 1010	90	5A	Z	0111 1010	122	7A	z
0011 1011	59	3B	;	0101 1011	91	5B	[0111 1011	123	7B	{
0011 1100	60	3C	<	0101 1100	92	5C	\	0111 1100	124	7C	

0011 1101	61	3D	=	0101 1101	93	5D]	0111 1101	125	7D	}
0011 1110	62	3E	>	0101 1110	94	5E	^	0111 1110	126	7E	~
0011 1111	63	3F	?	0101 1111	95	5F	_	0111 1111	127	7F	DEL

3.2 Засоби обробки інформації

При сучасному розвиненні програмного забезпечення існує безліч різноманітних програмних засобів обробки [інформації](#), написаних різними мовами [програмування](#). Це різноманіття пов'язано із специфікою кожної галузі, в якій проводиться обробка. Наприклад при обробці графічних зображень широко використовуються методи розпізнавання образів, криптографічні методи. Що базуються на перетворенні Фур'є тощо.

Чи не кожна фірма може дозволити собі замовити у розробника програму, що дозволить ефективно обробляти інформацію, пов'язану саме зі сферою діяльності цієї фірми. Такий підхід є навіть бажаним, оскільки автоматизовані системи обробки базуються на визначеній базі даних, структура якої може суттєво відрізнятися у різних підприємств, не кажучи вже про різні галузі. Одним із найрозповсюджених засобів обробки інформації є пакет [Microsoft Office](#), оскільки він встановлений майже на кожному комп'ютері. Його діапазон можливостей досить широкий, проте примітивний, якщо користувач не може безпосередньо працювати у програмному середовищі, на якому розроблений офіс. Серед засобів, доступних широкому класу споживачів є організація баз даних, відповідно виконання запитів та пошуку інформації, фільтрування інформації, графічне представлення тощо.

Для обробки графічних зображень (відсканованих) використовується Fine Reader, що є чудовим прикладом розпізнавання інформації. Для роботи із звуковими файлами використовують аудіо редактори. Наприклад Sound Forge. Для специфічного та глибокого аналізу статистичних даних використовується пакет SPSS, який розроблено спеціально для обробки даних із застосуванням статистичних методів (перевірка гіпотез, графічне зображення тощо).

В теперішній час внаслідок глобального поширення комп'ютерних систем в галузі автоматизації промислових процесів все частіше застосовуються системи збору даних і оперативного диспетчерського управління (SCADA – Supervisory Control And Data Acquisition System). [SCADA](#) – це тільки один з компонентів автоматизованих систем управління, які на сучасному етапі є складним комплексом програмних і апаратних засобів. Переважна більшість автоматизованих систем управління будується на базі промислових контролерів, які є первинними засобами збору, обробки інформації, регулювання технологічними параметрами, аварійної сигналізації, захисту і блокування (нижній рівень системи). Оброблена контролерами інформація передається до комп'ютеризованих систем, які є робочим місцем оператора-технолога, де відбувається подальша обробка даних процесу і представлення оператору в інтуїтивно зрозумілому вигляді. SCADA-системи в ієрархії програмно-апаратних засобів промислової автоматизації знаходяться на верхньому рівні. Якщо спробувати стисло охарактеризувати основні функції, то можна сказати, що SCADA-система збирає інформацію про технологічний процес, забезпечує інтерфейс з оператором, зберігає історію процесу і здійснює управління процесом в тому обсязі, в якому це необхідно.

На даному етапі все більшого розвитку також набувають методи [людино-орієнтованої комп'ютерної обробки даних](#).

Аналітико-синтетична обробка документів. Поняття «аналітико-синтетична обробка документів» має різне тлумачення і різне наповнення, що відображається на змісті навчальної дисципліни. Спираючись на сучасні термінологічні словники, [стандарти](#), посібники, можна встановити, що в інформатиці під аналітико-синтетичною обробкою інформації розуміють переробку інформації, що міститься в документах з метою вилучення необхідних відомостей, їх оцінки, порівняння і узагальнення. В [бібліографії](#) аналітико-синтетичну обробку документів розуміють як процес бібліографування, а саме процеси, що забезпечують ідентифікацію

документів, інформацію про них. В каталогізації метою аналітико-синтетичної обробки є розкриття складу і змісту документального фонду, що дає можливість ідентифікувати документи і здійснювати бібліографічний пошук. Кожна галузь визначає певні види або процеси аналітико-синтетичної обробки. В інформатиці це, як правило, [анотування](#), [реферування](#), вилучення фактів і підготовка оглядів; в бібліографії – складання бібліографічних описів (БО), анотування, індексування, складання бібліографічних оглядів та покажчиків; в каталогізації – складання БО, індексування, анотування.

Статистична обробка інформації. Обробку інформації можна здійснити, використовуючи комплекси засобів обчислювальної та інформаційної техніки, за допомогою якої раціонально виконуються чотири групи основних операцій: збір і передача інформації в ЕОМ, її обробка за заздалегідь розробленими алгоритмами та програмами, видавання обробленої інформації користувачам, зберігання й пошук інформації. Залежно від експлуатаційних можливостей і ступеня застосування зазначеної техніки при виконанні цих операцій відбувається часткова й комплексна механізація операцій, часткова й повна їх автоматизація. Повна чи комплексна автоматизація інформаційних операцій на практиці й у літературі відома як система машинної чи автоматизованої обробки інформації.

У сучасних умовах, коли потрібно не лише збирати й обробляти інформацію, а й мати змогу забезпечувати нею менеджерів і фахівців різних рівнів, зазначена система дістала подальший розвиток і діє як система автоматизованого збирання та обробки інформації. Щоб організувати систему автоматизованого збирання та обробки інформації, яка використовується для управління об'єктом, потрібно створити такі процеси, за яких вірогідна первинна інформація (яка характеризує виробничо-господарську та іншу діяльність) один раз у мінімальному складі в ритмі виробництва вимірювалася б і фіксувалася (а іноді й первинно оброблялася) й у такому самому ритмі (при потребі) передавалася каналами зв'язку до ЕОМ користувачів і комплексно (системно, тобто всебічно за відповідними алгоритмами) оброблялася (разом з

умовно-постійною) на ЕОМ так, щоб одержана в максимумі результатна інформація була закінчена обробкою і повністю відповідала б усім вимогам управління об'єктом, у тому числі інформуванню менеджерів і фахівців різних рівнів, а також процесам конструкторської і технологічної підготовки виробництва, нормування й планування, обліку й контролю, складання установленої зведеної звітності й проведення комплексного економічного аналізу за різні періоди часу та по різних структурних ланках і об'єкту управління в цілому. В цьому й полягає її сутність.

Така система збирання та обробки інформації може створюватися на різних рівнях, у масштабі різних організаційно-управлінських ланок — цеху чи виробничої одиниці, підприємства чи виробничого об'єднання, галузі, регіону тощо. Зі збільшенням масштабності управлінської ланки значно зростають трудомісткість і складність організації системної обробки інформації.

Системна обробка інформації. При системі автоматизованого збирання й обробки інформації, яка діє в ритмі виробництва на об'єкті управління, не лише збирають і обробляють у ритмі виробництва всі первинні дані, що характеризують виробничо-господарську та іншу діяльність, а й забезпечують (через систему інформування) результатною (обробленою) інформацією всіх менеджерів і фахівців різних рівнів. Завдяки безпаперовій технології користувачі оперативно одержують на екрані ЕОМ об'єктивно необхідну інформацію згідно з делегованими повноваженнями, якими вони наділені менеджерами вищого рівня, функціональними обов'язками, посадовими правами та мірою відповідальності. Вихідна інформація, як правило, є повністю оброблена, включає результати елементарного аналізу даних і є необхідною для вироблення та прийняття управлінських рішень.

Усе це звільняє менеджерів і фахівців різних рівнів від рутинної праці зі збирання, зберігання й обробки даних та їх елементарного аналізу і дає змогу зосередитися на творчій праці та контролі за виконанням управлінських рішень. Розробка і застосування нових методів управління та вдосконалення

діючих значно поліпшує якість і оперативність управління, його дієвість, а також забезпечує своєчасне вироблення й прийняття управлінських рішень. Наприклад, при цій системі є можливість використати метод управління за відхиленнями, завдяки якому значно зменшуються обсяги аналізованої інформації — до користувача надходить на екран його ЕОМ лише релевантна суттєва інформація в виді відхилень, на основі якої виробляється та приймається відповідне управлінське рішення. Дублююча та зайва (яка не використовується в даному разі) інформація на екран ЕОМ не виноситься, оскільки вона відкидається. Високі експлуатаційні можливості сучасної обчислювальної техніки і засобів комунікації дозволяють усебічно й на всіх рівнях оперативно збирати й обробляти інформацію виробничого, комерційного та іншого характеру. А це означає, що не потрібні проміжні управлінські структури в загальній організаційній структурі об'єкта, наприклад, цехи. Згадана інформаційна система спроможна не лише розробляти оптимальні планові та інші показники, а й доводити їх до дільниць, бригад, конкретних робочих місць, а також у ритмі виробництва збирати фактичні дані, які характеризують виконані операції, процеси тощо. В цьому випадку проміжній структурі — цеху, з функціями, які він виконував раніше, не буде місця. Замість нього діятиме ланка менеджерів середнього рівня з контрольними функціями. Зазначимо також, що як показала практика, завдяки застосуванню методів автоматизованого збирання в ритмі виробництва первинної фактичної інформації на робочих місцях її масового виникнення, є змога знизити на 80—85% ручну працю на цих процесах. Ця інформація стає єдиною як для обліку, так і для управлінських цілей. Це дає змогу об'єднати три види діючого обліку (оперативно-технічний, статистичний і бухгалтерський) в єдиний. Первинні дані цього обліку є вірогідними, документованими і об'єктивними. На їх основі автоматично складають установлену для об'єкта управління зведену звітність, а потім передають дані цієї зведеної звітності вищим за субординацією установам каналами зв'язку. Застосовуючи зазначені методи формування первинної

інформації на ЕОМ, вводять у пам'ять цієї машини мінімум інформації, а одержують (тобто виводять після обробки на ЕОМ) максимум, що значно скорочує затрати праці. Завдяки використанню ЕОМ і нових інформаційних технологій значно вдосконалюються форми єдиного обліку за рахунок спрощення їх, виведення з них дублюючої, розрахункової і підсумкової інформації. Таким чином, системне оперативне збирання та обробка інформації серйозно впливає на методи управління соціально-економічним об'єктом.

Автоматизована обробка інформації. Експлуатаційні можливості сучасного комплексу технічних засобів, що використовується в системі автоматизованого збирання й обробки інформації, дають змогу автоматизовано виконувати цілу низку процедур у цих функціях. Стан науково-практичних розробок та технічний рівень згаданого комплексу визначили можливості автоматизованого виконання таких процедур управлінського процесу:

- у прогнозуванні та плануванні — багатоваріантні розрахунки під час розробки прогнозів, перспективних і поточних економічних та соціальних планів розвитку підприємства, а також оперативно-виробничих планів і планів з технічної підготовки виробництва з метою подальшого визначення оптимальних взаємозв'язаних наборів показників планування в погодинному (година, зміна, тиждень тощо) і в пооб'єктному (робоче місце, ділянка тощо) аспектах;
- в організації — моделювання організаційних структур управління та імітація процесів виробництва при різних критеріях і параметрах з метою вибору оптимальних;
- при координації і регулюванні — подання команд на робочі місця (поки що на низовому рівні управління виробництвом) відповідно до плану, технологічного процесу чи інструкції, складених на ті чи інші види робіт або операції;

- у контролі — спостереження за станом керованого об'єкта за всіма параметрами, а також за своєчасним і повним виконанням керівних команд;
- в обліку — одноразове збирання (в ритмі виробництва) й системна обробка всієї фактичної (разом з довідковою, плановою, нормативною та іншою) вірогідної інформації про наявність та рух ресурсів, а також про стан, процеси та явища, що мають місце у виробничо-господарській та іншій діяльності підприємства;
- в аналізі — зіставлення нормативних, планових і фактичних показників, що характеризують ті чи ті операції або процеси виробничо-господарської та іншої діяльності, виявлення відхилень (у кількісних, вартісних, відносних та інших величинах) від заданих параметрів із зазначенням причин і винуватців цих відхилень, оцінка виконання плану в різних аспектах та виявлення факторів, що впливають на ці відхилення;
- у звітності — автоматичне формування (на основі первинних даних) зведених показників для типових форм установленої бухгалтерської, статистичної та іншої звітностей за допомогою спеціальних перевідних масивів — довідників, — а також одночасне створення машинних носіїв зі зведеними показниками звітності для передавання каналами зв'язку їх до зовнішніх установ (інституцій) вищого рівня.

Відомо, що в умовах функціонування системи автоматизованого збирання й обробки економічної інформації первинну базу наукового прогнозування і планування становлять обґрунтовані норми й нормативи, які застосовуються в усіх процесах та явищах, що мають місце на об'єкті управління. Процедури їх розробки, аналізу та всебічного використання також повністю автоматизовані.

Експлуатаційні можливості, наприклад, ЕОМ, що використовуються як АРМ фахівців різних рівнів, дають змогу оперативно (в ритмі виробництва) й системно одноразово збирати (в мінімальному складі) вірогідну первинну масову інформацію на робочих місцях та ділянках, від комплексних бригад та інших ланок виробництва, де вона масово виникає; потім усю цю

інформацію попередньо обробляти таким чином, щоб у ритмі виробництва повністю забезпечувати (в максимальному складі) при безпаперовій інформаційній технології об'єктивно необхідною вихідною (результатною) інформацією, з одного боку, широкі потреби користувачів — менеджерів та фахівців даного (низового) рівня управління, — а з другого, передавати в стислому чи іншому вигляді цю інформацію на вищий (середній та верхній) рівень управління. На цьому рівні зібрана інформація (після додаткової обробки, а часом і в такому, первинному вигляді) використовується не лише для аналізу й прийняття управлінських рішень, а й для обліку, контролю, комплексного економічного аналізу результатів виробничо-господарської та іншої діяльності структурних ланок і об'єкта управління в цілому, а також для складання установленної зведеної бухгалтерської, статистичної звітностей.

Повне оперативне інформування користувачів ([менеджерів](#)) різних рівнів про процеси та явища, що відбуваються на об'єктах управління, створює умови для виконання в повному (в інформаційному аспекті) обов'язі посадових функцій відповідно до наданих прав, обов'язків та міри відповідальності за доручену справу. Ідеться про [делегування повноважень](#) вищим менеджером нижчому, тобто про передачу завдань, прав і обов'язків особі, яка бере на себе відповідальність за їх повне і неухильне виконання. При цьому повноваження, звісно, певною мірою обмежують права на використання тих чи тих ресурсів, а також дозволяють спрямовувати зусилля лише на виконання визначених завдань (тобто делегованих прав).

3.3 Технології обробки інформації

3.3.1 Основна мета та історія розвитку

В ході інформаційного процесу інформація, що циркулює на підприємстві або в організації, піддається тій чи іншій обробці в залежності від роду їх діяльності. За місцем виникнення виділяють вхідну і вихідну, внутрішню і зовнішню інформацію. В процесі обробки інформація може бути первинною і

вторинною, проміжною і результативною, при цьому оброблювані дані перетворюються з одного виду в інший. За час розвитку інформаційного суспільства трудовитрати на обробку даних зростають і вимагають вдосконалення застосовуваних технологій.

Технологія (гр. *Techne* - майстерність, *logos* - вчення, вчення про майстерність) - сукупність знань про способи і засоби виробничих процесів, при яких відбувається необхідне якісна зміна оброблюваних об'єктів.

Інформаційна технологія - процес, що використовує сукупність засобів і методів збору, обробки і передачі даних для отримання інформації нової якості про стан об'єкта, процесу або явища.

Мета інформаційної технології - виробництво інформації для її аналізу людиною і подальшого прийняття рішень по здійсненню будь-яких дій. У більш вузькому розумінні інформаційна технологія являє собою сукупність чітко визначених цілеспрямованих дій людини по переробці інформації на комп'ютері. Технологічний процес переробки інформації складається з етапів, операцій і конкретних дій оператора, що виконує обробку даних. У структурі можливих операцій з даними можна виділити наступні:

- збір даних та їх формалізація, тобто приведення до однакової форми;
- фільтрація і сортування;
- обробка і перетворення даних відповідно до поставленим завданням;
- архівація даних, тобто організація зберігання даних в компактній, зручній і доступній формі;
- захист даних - комплекс заходів, спрямованих на запобігання втрати даних і їх модифікації;
- транспортування даних, тобто прийом і передача даних між віддаленими учасниками інформаційного процесу.

Історія розвитку інформаційних технологій включає кілька етапів, пов'язаних з кардинальними змінами в сфері обробки інформації.

Перший етап пов'язаний з винаходом писемності. Засобами збору, зберігання і обробки інформації тут служили перо, чорнило, папір і книги,

ефективність інформаційної обробки на цьому етапі була вкрай низькою. Винахід друкарства в середині XVI ст. значно підвищило ефективність обробки інформації, виникли такі кошти, як складальна дошка і друкарський верстат.

На зміну "ручний технології" в кінці XIX ст., з появою телеграфу, телефону, радіо, прийшла "механічна" технологія, що дозволяє оперативно передавати інформацію.

Створення електричних друкарських машинок, телебачення, копіювальних апаратів, магнітофонів до середини XX в. призвело до виникнення "електричних" інформаційних технологій.

З другої половини XX в. і з появою ЕОМ, а потім персонального комп'ютера почався новий етап у розвитку інформаційних технологій - "електронні" технології.

Електронна обчислювальна машина - універсальний пристрій введення, виведення, накопичення, обробки та передачі інформації для вирішення обчислювальних та інформаційних задач. Термін "комп'ютер" вживається в тому ж сенсі, що і термін "ЕОМ". ЕОМ - електронна машина, так як складається з електронних схем, і обчислювальна машина, так як обробляє інформацію в цифровій формі, виконуючи обчислення, чисельні арифметичні і логічні операції без втручання людини. Цифрова форма подання будь-яких даних забезпечує комп'ютеру такі властивості, як універсальність, придатність для вирішення різноманітних завдань.

Вперше проект аналітичної машини (обчислювального автомата) в складі пристрої введення, пристрої пам'яті, процесора, пристрої виведення був запропонований в XIX в. Чарльзом Бебіджем. Він же вперше висунув ідею програмного управління такою машиною. Подальший розвиток цієї ідеї знайшло своє продовження при побудові перших електронно-обчислювальних машин. Функціонування ЕОМ базувалося на двійковій системі числення для подання чисел і розміщення програми управління в пристрої. Перші ЕОМ розроблялися в 1943-1947 рр. в США і Англії, в континентальній Європі перша

"мала електронна рахункова машина" (МЕСМ) була створена в СРСР в 1948-1951 рр.

Види ЕОМ. Електронно-обчислювальні машини прийнято класифікувати за деякими ознаками.

За фізичним поданням оброблюваної інформації виділяють:

- аналогові обчислювальні машини безперервного дії, які працюють з інформацією, представленою в безперервній (аналоговій) формі, тобто у вигляді безперервного ряду значень будь-якої фізичної величини (найчастіше електричного напруги);

- цифрові обчислювальні машини, які працюють з інформацією в дискретній формі (цифровий);

- гібридні обчислювальні машини комбінованої дії, що поєднують в собі переваги аналогових і цифрових обчислювальних машин і використовуються для вирішення завдань управління складними швидкодіючими технічними комплексами.

По етапах створення ЕОМ виділяють кілька поколінь розвитку комп'ютерної техніки, які формувалися протягом ХХ в.

До першого покоління відносять машини, створені в 1950-і рр. на основі електронних ламп. В цей час були розроблені вітчизняні машини: МЕСМ (мала електронна рахункова машина), БЕСМ (велика електроннолічильна машина), "Стріла", серія "Урал", М-20. Основним застосуванням перших ЕОМ було виконання науково-технічних розрахунків.

Через десятиліття з'явилися ЕОМ, створені на дискретних напівпровідникових приладах (транзисторах). Друге покоління ЕОМ застосовувалося для технічних і економічних розрахунків.

Машини третього покоління з'явилися в 1970-і рр. і були розроблені на напівпровідникових інтегральних схемах з малої і середньої ступенем інтеграції (сотні, тисячі транзисторів в одному корпусі). Це покоління ЕОМ початок застосовуватися в управлінні і проведенні економічних розрахунків.

Четверте покоління ЕОМ сформувалося в 1980-і рр. на базі великих і надвеликих інтегральних схем - мікропроцесорів (десятки тисяч - мільйони транзисторів в одному кристалі). Метою ЕОМ цього покоління вже було подання інформації та більш широке використання в управлінні.

Так, 1990-ті рр. характеризуються створенням ЕОМ з багатьма десятками паралельно працюючих мікропроцесорів, що дозволяють будувати ефективні системи обробки знань. Для цього покоління характерні застосування персональних комп'ютерів, телекомунікаційна обробка даних, комп'ютерні мережі, широке застосування систем управління базами даних, елементи інтелектуальної поведінки систем обробки даних і пристроїв.

Створення оптоелектронних ЕОМ з масовим паралелізмом і нейронної структурою відноситься до початку ХХІ ст. Передбачається, що в комп'ютерах наступного покоління відбудеться якісний перехід від обробки даних до обробки знань.

3.3.2 Інформаційні технології в умовах функціонування систем обробки інформації

Технологія інформації характеризується концентрацією інформації та обчислювальних засобів, обробкою значних обсягів інформації в неоперативному режимі, застосуванням середніх і великих ЕОМ, які встановлені на обчислювальних центрах (ОЦ). В умовах централізованих автоматизованих інформаційних системах реалізується хост-орієнтована технологія: усі обчислювальні потужності зосереджені на великій ЕОМ (мейн-фреймі) або міні-ЕОМ (хості), а термінали, через які користувачі підключаються до хосту, не можуть самостійно робити обчислення. Участь користувача в технології обробки зводиться до представлення і контролю початкових даних, аналізу результатів, участі в розробці постановок задач. Централізовані автоматизовані інформаційні системи функціонують на основі інтегрованої бази даних.

Перевагою централізованих технологій є максимальне завантаження ЕОМ. Але при цьому залишається велика частка ручних операцій у технологічному процесі, складна організація діалогового режиму розв'язання задач, не досягається неперервність обробки даних.

Технологія в умовах децентралізованих систем визначається обробкою інформації в місцях її виникнення та оброблення на ПК. В умовах децентралізованих автоматизованих інформаційних системах реалізується два типи моделей обчислювального процесу: модель процесу з розподіленими ресурсами і модель технології "клієнт – сервер".

Технологія, яка реалізує модель процесу з розподіленими ресурсами, реалізує всі обчислення на ПК і в мережі. Вона дозволяє всім користувачам мережі розділяти коштовні ресурси: принтери, дискові накопичувачі, модеми тощо.

Технологія "клієнт - сервер" дозволяє розподіляти процес обчислення між ПК ("клієнтом") і більш потужною ЕОМ, яка використовується багатьма користувачами ("сервером"). Технологія "клієнт–сервер" заснована на обміні електронними повідомленнями (клієнт посилає запит на сервер і отримує відповідь) і керується подіями (при появі конкретної події на ньому реагує спеціальний триггер, який генерує і відправляє конкретне повідомлення, наприклад, інформує про зміну даних).

Технологічний процес у децентралізованих автоматизованих інформаційних системах характеризується участю користувача у формуванні БД і в розв'язанні задач управління. При цьому користувач спілкується з ПК у діалоговому режимі. У децентралізованих автоматизованих інформаційних системах застосовується розподілена БД. У результаті запровадження децентралізованих технологій підвищується оперативність управління, скорочується трудомісткість обробки інформації, користувачем забезпечується контроль вхідних даних і результатів розрахунку.

3.3.3 Основні режими обробки інформації

Незалежно від конкретного змісту процесу обробки інформації користувач взаємодіє з комп'ютером в одному з двох режимів: пакетному або діалоговому.

Пакетний режим – це режим, при якому користувач не має доступу до машинних ресурсів, а обробка інформації виконується у спеціальному підрозділі, основною функцією якого є обробка інформації.

Через відсутність у користувача змоги безпосереднього доступу до машинних ресурсів існують затримки у видачі результатів, що пов'язано з особливостями технології обробки інформації і необхідністю обробки багатьох завдань в одному місці почергово. Внаслідок можливих затримок не розв'язуються деякі задачі оперативного характеру (задачі, які потребують видачі результату з непередбаченою періодичністю та в короткий термін).

Найчастіше оперативна результатна інформація вже наступного дня або години не потрібна, тому затримка з видачею результатів обробки в пакетному режимі призводить до того, що в такому режимі розв'язуються лише задачі регламентного характеру. У регламентних задач відома періодичність їх розв'язування та термін, до якого необхідно отримати результат. Наприклад, заробітна плата нараховується один або два рази на місяць, результат розрахунку необхідно передати до бухгалтерії за один-два дні до видачі грошей із каси.

Бувають випадки, коли на робочому місці користувача встановлено ПК, але обробка інформації здійснюється у пакетному режимі. Програми обробки інформації побудовані так, що кожна програма запускається в роботу окремою командою, без втручання користувача виконує всі розрахунки і формує результати. До функцій користувача належить лише набрати на клавіатурі ЕОМ команди для запуску програм, тобто користувач працює як оператор. Цей варіант обробки інформації можна віднести до прорахунків у проектних розробках, оскільки для користувача не створений зручний інтерфейс для роботи з ЕОМ.

Пакетний режим обробки даних передбачає формування низки завдань (програм) в єдиний пакет з його подальшим виконанням без безпосередньої участі користувача. Хронологічно період широкого застосування пакетної технології відповідав централізованій обробки в обчислювальних центрах (обчислювальних центрах колективного користування - ОЦКП). При цьому користувач багаторазово взаємодіяв з оператором: віддавав своє завдання, отримував роздруковане завдання, як правило, з виявленими помилками, усував помилки і знову віддавав оператору завдання і так далі до отримання прийнятних результатів обробки даних. Головним недоліком пакетної технології є неможливість оперативного внесення змін у процес обробки в ході його виконання. В даний час завдяки широкому розповсюдженню персональних комп'ютерів ступінь використання пакетної технології значно зменшилася. Фактично пакетний режим реалізується при початковій завантаженні комп'ютера, коли послідовно в заданому порядку запускаються і виконуються наперед задані послідовності програм. Для користувача є можливість сформувати власні набори послідовно запускаються і виконуваних програм за допомогою підготовки і подальшого виконання, що їх містять назви командних файлів. Крім того, риси пакетної технології проглядаються при роботі з електронною поштою, коли послідовно в автоматичному режимі здійснюється прийом пошти, перевірка її на наявність вірусів, відправка своїх листів.

Збір та реєстрація інформації, введення та обробка не збігаються за часом. Спочатку користувач збирає інформацію, формуючи її в пакети згідно з видом завдань або якоюсь іншою ознакою. Після завершення прийому інформації проводиться її введення та обробка, тобто, відбувається затримка обробки. Цей режим використовується, як правило, при централізованому способі обробки інформації.

Діалоговий режим обробки даних означає обмін повідомленнями між користувачем і системою в реальному часі, тобто в темпі реакції користувача на події, що створює природні умови для ефективного управління процесом

обробки. При роботі на багатокористувацьких системах (високопродуктивних комп'ютерах) діалогова технологія реалізується в режимі поділу часу, коли процесорний час виділяється всім користувачам (важливість справ) періодично малими квантами часу. У кожного користувача при цьому створюється враження безперервності процесу обробки його даних.

Інтерактивний режим роботи ЕОМ – це режим, в якому користувач має безпосередній доступ до машинних ресурсів і впливає на процес обробки інформації. Такий вплив спричиняє відповідну реакцію системи.

Діалоговий режим роботи ЕОМ – це режим, в якому людина і ЕОМ обмінюються інформацією в темпі, який можна порівняти з темпом обробки інформації людиною.

Інтерактивному і діалоговому режиму притаманна робота у вигляді діалогу. Користувач безпосередньо бере участь у реалізації процесу обробки даних на ЕОМ, тобто посередник в особі різних підрозділів обчислювального центру відсутній. З'являється можливість втручання користувачів у процес розв'язування задач на ЕОМ, що особливо важливо для тих задач, які реалізуються операційними методами і в алгоритмі розв'язання яких передбачаються формалізовані параметри (вони заздалегідь не можуть бути введені в ІБ, а задаються користувачем при одержанні результатів розв'язання задачі на ЕОМ). Підвищується культура управління. Діалоговий режим може використовуватися як при централізованій, так і при розподіленій обробці інформації.

Виділення діалогового режиму обумовлено не лише особливостями доступу до ресурсів ЕОМ, але і принципами побудови програмного забезпечення для обробки інформації. При діалоговому режимі програми побудовані таким чином, що користувач може вибирати під час діалогу з ЕОМ той чи інший розрахунок або визначати напрямки подальшої роботи.

Діалоговий режим передбачає багаторазове надходження запитів користувача через дисплей у систему. На кожен запит дається відповідь або зустрічне запитання. При цьому запит формується системою автоматично, на

основі здійснення діалогу між ЕОМ і користувачем. У процесі діалогу ЕОМ задає користувачеві запитання, супроводжуючи їх необхідними підказками (коментарями).

Вхідна інформація зберігається в БД, алгоритм і програми заздалегідь введені в ЕОМ. Відповідь видається на принтер, на екран або магнітні носії інформації.

Застосування діалогової технології дозволило вести обробку інформації в режимі реального часу, коли результати обчислень видаються в необхідні моменти фактичного проходження виробничого процесу. Інформація видається кінцевому користувачеві в наочному вигляді, і це дозволяє підвищити оперативність і достовірність обробки даних.

Діалоговий режим вимагає певного рівня технічної оснащеності користувача, тобто наявності терміналу або ЕОМ, пов'язаних центральною обчислювальною системою каналами зв'язку. Цей режим використовується для доступу до інформації, обчислювальним або програмних ресурсів. Можливість роботи в діалоговому режимі може бути обмежена в часі початку і кінця роботи, а може бути і необмеженою. Широке поширення діалогової технології було об'єктивно обумовлено рядом причин:

- наявністю широкого кола завдань, що припускають пошук рішення в процесі обробки даних при безпосередній участі користувача;
- появою на робочих місцях користувачів персональних комп'ютерів з дружнім інтерфейсом, які мають ефективні засоби підтримки інтерактивної взаємодії з користувачем (миша, кольорові дисплеї з високим дозволом і підтримкою графіки, вінчестери великої ємності, аудіопристрої, операційні системи розвиненим віконним інтерфейсом та ін.) при їх мінімальному рівні комп'ютерної грамотності;
- розвитком локальних обчислювальних мереж та засобів телекомунікацій, що забезпечують взаємодію користувачів на будь-якому видаленні один від одного.

Діалогова технологія забезпечує ряд переваг порівняно з пакетною технологією:

- можливість налаштування ресурсів комп'ютера під індивідуальні запити користувача з метою оптимізації ходу виконання конкретного завдання;
- забезпечення умов для оперативного внесення змін у процес обробки даних на основі аналізу проміжних результатів (використання експертних систем);
- паралельна робота ряду користувачів над спільним завданням (ведення окремих ділянок обліку різними бухгалтерами з формуванням зведених документів) з використанням можливостей мережевих технологій;
- ефективне використання екранних підказок і довідкової системи;
- оперативний пошук за необхідності інформаційних ресурсів в глобальних мережах;
- можливість швидкого перетворення форм відображення даних і результатів їх обробки (у вигляді таблиць, графіків, діаграм та ін);
- прискорення процесів налагодження використовуваного прикладного програмного забезпечення;
- можливість візуального контролю (використання шаблонів, довідників для введення типових даних, введення даних за зразком та ін) як на етапі введення даних, так і на етапі вирішення;
- оперативне управління контролем доступу до інформаційних ресурсів з урахуванням рівнів привілеїв користувачів.

Режим реального масштабу часу означає здатність обчислювальної системи взаємодіяти з контрольованими або керованими процесами в темпі протікання цих процесів. Час реакції ЕОМ повинно задовольняти темпу контрольованого процесу або вимогам користувачів і мати мінімальну затримку. Як правило, цей режим використовується при децентралізованій і розподіленій обробки даних.

Режим телеобробки дає можливість віддаленому користувачеві взаємодіяти з обчислювальною системою.

Інтерактивний режим передбачає можливість двосторонньої взаємодії користувача з системою, тобто у користувача є можливість впливу на процес обробки даних.

Режим поділу часу припускає спроможність системи виділяти свої ресурси групі користувачів по черзі. Обчислювальна система настільки швидко обслуговує кожного користувача, що створюється враження одночасної роботи декількох користувачів. Така можливість досягається за рахунок відповідного програмного забезпечення.

Однопрограмний і багатопрограмний режими характеризують можливість системи працювати одночасно по одній або декількох програмах.

Регламентний режим характеризується визначеністю в часі окремих завдань користувача. Наприклад, отримання результатних зведень по закінченні місяця, розрахунок відомостей нарахування зарплати до певних дат і т.д. Терміни рішення встановлюються заздалегідь за регламентом на протигагу до довільним запитам.

3.3.4 Комплекс технічних засобів обробки інформації

Інформаційні технології використовують програмні та технічні засоби. Комплекс технічних і програмних засобів постійно розширюється, що обумовлено розвитком ІС у бік застосування різних інформаційних середовищ, включаючи мультимедіа.

Програмні засоби інформаційних технологій забезпечують можливість програмної обробки інформації в електронному вигляді вони включають:

- системне програмне забезпечення, яке забезпечує роботу електронно-обчислювальної техніки та комп'ютерних мереж;.
- прикладне програмне забезпечення, яке дозволяє розв'язувати конкретні управлінські завдання та опрацьовувати інформацію відповідно до

поставлених перед працівником завдань та його безпосередніх посадових обов'язків.

Технічні засоби обробки інформації діляться на дві великі групи: основні і допоміжні засоби обробки.

Допоміжні засоби - це обладнання, що забезпечує працездатність основних засобів, а також обладнання, що полегшує і робить управлінську працю комфортніше. До допоміжних засобів обробки інформації належать засоби оргтехніки та ремонтно-профілактичні засоби. Оргтехніка представлена досить широкою номенклатурою засобів, від канцелярських товарів, до засобів доставляння, розмноження, зберігання, пошуку і знищення основних даних, засобів адміністративно виробничої зв'язку і так далі, що робить роботу управлінця зручною і комфортною. Основні засоби - це знаряддя праці з автоматизованої обробки інформації. Відомо, що для управління тими чи іншими процесами необхідна певна управлінська інформація, що характеризує стан і параметри технологічних процесів, кількісні, вартісні і трудові показники виробництва, постачання, збуту, фінансової діяльності.

До **основних засобів** технічної обробки відносяться: засоби реєстрації та збору інформації, засоби приймання та передачі даних, засоби підготовки даних, засоби введення, засоби обробки інформації та засоби відображення інформації.

Технічні засоби для інформаційних систем поділяються на класи:

- засоби збору і реєстрації інформації (персональні комп'ютери, сканери, автоматичні датчики);
- засоби передавання інформації (локальні, регіональні та глобальні обчислювальні мережі, Інтернет);
- засоби зберігання даних (сервери баз даних, файлові сервери, локальні комп'ютери, магнітні диски, оптичні (лазерні) диски, цифрові відеодиски);
- засоби обробки даних (мікрокомп'ютери, міні-комп'ютери, великі комп'ютери);

- засоби виведення інформації (відеомонітори, принтери, графо-побудовники).

В залежності від використання технічних засобів розрізняють **способи обробки облікової інформації** - сукупність методики та технічних засобів, за допомогою яких здійснюється обробка облікової інформації (рис.3.1).



Рисунок 3.1 - Способи обробки облікової інформації

Переваги безкомп'ютерного способу обробки облікової інформації:

- незначні вимоги до кваліфікації облікових працівників;
- складність випадкової втрати облікової інформації;
- низькі витрати на утримання облікового апарату та системи обліку підприємства.

Недоліки безкомп'ютерного способу обробки облікової інформації:

- значна тривалість та трудомісткість облікового процесу;
- складність випадкової втрати облікової інформації;
- значні штати облікового апарату;
- складність здійснення контролю;
- низький рівень деталізації облікової інформації;
- значні обсяги рутинних операцій.

Перевагами комп'ютеризованого способу обробки інформації є:

- можливість широкого розподілу праці між обліковими працівниками;
- можливість кільком бухгалтерам працювати з одним регістром;
- можливість значної деталізації даних в аналітичному обліку;

- поєднання документування і запису в обліковій реєстрі;
- автоматичний контроль за арифметичними підрахунками;
- значна швидкість одержання результатної інформації;
- можливість ведення багатовимірного і багаторівневого обліку;
- скорочення потоку паперових документів і перехід до безпаперової технології;
- можливість щоденного формування звітності підприємства за потребою;
- можливість вибору форми подання інформації;
- можливість пошуку інформації на своїх робочих місцях з метою економії робочого часу.

Недоліками **комп'ютеризованого** способу обробки інформації є:

- недостатнє вирішення проблеми захисту інформації від несанкціонованого втручання;
- проблема юридичної доказовості складених електронних первинних документів;
- можливість втрати чи псування інформації при відключенні електроенергії, проникненні комп'ютерних вірусів тощо;
- значні одноразові витрати на здійснення автоматизації, придбання комп'ютерної техніки, програмного забезпечення та навчання облікового персоналу.

Механізований спосіб обробки облікової інформації дозволяє за мінімальних витрат отримати значне зниження трудомісткості облікових операцій та значно підвищити продуктивність праці облікового персоналу. Незважаючи на суттєві переваги, він підвищує імовірність одержання помилок при виконанні операцій обміну обліковою інформацією.

4 ЗБЕРІГАННЯ ІНФОРМАЦІЇ

4.1 Збір та зберігання інформації

Останні два десятиріччя масове виробництво персональних комп'ютерів і стрімке зростання Інтернету суттєво прискорили становлення інформаційного суспільства в розвинених країнах світу.

В інформаційному суспільстві головним ресурсом є інформація, саме на основі володіння інформацією про всілякі процеси і явища можна ефективно й оптимально будувати будь-яку діяльність. Більша частина населення в інформаційному суспільстві зайнята в сфері обробки інформації або використовує інформаційні й комунікаційні технології у своїй повсякденній виробничій діяльності.

Під **інформацією** розуміють корисний зміст явища факту, вилученого з даних. Тому на практиці дуже часто неправомірно встановлюють знак рівності між даними та інформацією. Дані реєструють явище і події, що відбуваються, а інформація — результат переробки та аналізу даних. Інформація переважно носить порівняльний характер. Вона складається тільки з нових відомостей, які оцінюються користувачем як корисні знання. Із загального протоколу вилучаються тільки відомості, які потрібні для визначеного користувача.

На сьогоднішній день інформаційні технології активно впливають на повсякденну діяльність будь-якої організації і є невід'ємною складовою частиною інформаційної інфраструктури цієї організації. Інформаційні технології дозволяють досліджувати та взаємопов'язувати складові ділянки діяльності організації між собою. Вони розвиваються надзвичайно швидкими темпами і охоплюють все ширше сфери діяльності таким чином, що будь-яка конкурентоспроможна діяльність в майбутньому не може бути сформована без детального аналізу можливостей застосування інформаційних технологій.

У сфері інформаційних технологій важливим є забезпечення належного стану інформації та її матеріальних носіїв, тобто зберігання інформації.

Зберігання інформації — це комплекс [заходів](#), який спрямований на забезпечення збереження повноти і цілісності сформованих даних про певну інформацію, створення і підтримання належних умов для їх використання, а також запобігання несанкціонованому [доступу](#), поширенню і використанню.

Зберігання інформації є однією з дій з інформацією, яка необхідна для забезпечення життєдіяльності та безпеки людини. Люди зберігають інформацію на різних носіях і для зберігання інформації створюють бібліотеки та медіатеки. Збереженою, тобто представленою на носію інформацією може користуватися будь-яка людина. Інформацію, яка представлена на носії рисунком, числами або текстом, можна довго зберігати та передавати на великі відстані. Також інформацію необхідно зберігати для того, щоб вона могла використовуватися, причому багаторазово, адже зберігання інформації - це спосіб розповсюдження інформації у просторі і в часі. Спосіб зберігання інформації залежить від її носія (книга-бібліотека, картина-музей, фотографія-альбом). ЕОМ призначена для компактного зберігання інформації з можливістю швидкого доступу до неї. Інформаційна система - це сховище інформації, забезпечене процедурами введення, пошуку, розміщення та видачі інформації. Наявність таких процедур - головна особливість інформаційних систем, що відрізняють їх від простих збирань інформаційних матеріалів. Наприклад, особиста бібліотека, в якій може орієнтуватися тільки її власник, не є інформаційною системою. У публічних же бібліотеках порядок розміщення книг завжди строго визначений. Завдяки йому пошук і видача книг, а також розміщення нових надходжень представляє собою стандартні, формалізовані процедури.

Для довготривалого зберігання інформації, її накопичення та передавання з покоління в покоління необхідно мати можливість її зберігати не тільки в пам'яті людини. Для цього використовуються зовнішні носії інформації. Інформація, закодована за допомогою природних і формальних мов, а також інформація у формі зорових і звукових образів зберігається у пам'яті людини. Але для довготривалого зберігання інформації, її накопичення та передавання з покоління в покоління використовуються носії інформації.

За оцінками фахівців, обсяг інформації, що фіксується на різних носіях, перевищує один ексабайт на рік (10^{18} байт/рік). Приблизно 80% всієї цієї інформації зберігається у цифровій формі на магнітних і оптичних носіях і тільки 20% - на аналогових носіях (папір, магнітні стрічки, фото- і кіноплівці).

Якщо всю записану у 2018 році інформацію розподілити на всіх жителів планети, то кожній людині доведеться по 280 Мбайт, а для її зберігання потрібно 88 мільйонів жорстких магнітних дисків по 25 Гбайт.

Носій інформації – матеріальний об'єкт, який призначений для зберігання та передавання інформації.



У табл. 4.1 представлені види носіїв інформації. Матеріальна природа носіїв інформації може бути різною: молекули ДНК, які зберігають генетичну інформацію; папір, на якому зберігаються тексти та зображення; магнітна стрічка, на якій зберігається звукова інформація; фото- та кінострічки, на яких зберігається графічна інформація; мікросхеми пам'яті, магнітні та лазерні диски, на яких зберігаються програми і дані в комп'ютері. Усі комп'ютери потребують певного місця для запуску програм і збереження файлів. Існують різні типи пам'яті та пристроїв для зберігання даних, кожен з яких використовують з певною метою.

Носії інформації характеризуються **інформаційною ємністю**, тобто кількістю інформації, яку вони можуть зберігати. Найбільш інформаційно ємнісними є молекули ДНК, які мають дуже малий розмір і щільно упаковані. Це дозволяє зберігати велику кількість інформації (до 10^{21} бітів в 1см^3), що дає можливість організму розвиватися з однієї-єдиної клітини, що містить всю необхідну генетичну інформацію.





Сучасні мікросхеми пам'яті дозволяють зберігати в 1см^3 до 10^{10} біт інформації, але це в 100 мільярдів разів менше, ніж у ДНК. Можна сказати, що сучасні елементи поки що істотно поступаються біологічній еволюції.

Але, якщо порівнювати інформаційну ємність традиційних носіїв інформації (книг) і сучасних комп'ютерних носіїв, то прогрес очевидний. На кожному гнучкому магнітному диску може зберігатися книга обсягом близько 600 сторінок, а на жорсткому магнітному диску або DVD - ціла бібліотека, що включає десятки тисяч книг.

Таблиця 4.1 – Види носіїв інформації

Носії інформації		
		ДНК
		Магнітні та оптичні диски

Подовження таблиці 4.1

		Папір
		Мікросхеми пам'яті
		Магнітна стрічка
		Фото- і кіноплівка

Велике значення має надійність і довготривалість зберігання інформації.

Молекули ДНК – найбільш довготривалий носій інформації. Вони десятки тисяч років зберігають генетичну інформацію даного виду. Молекули

ДНК мають велику стійкість до можливих пошкоджень, оскільки існує механізм виявлення пошкоджень їх структури та самовідновлення.

У аналогових носіїв, надійність (стійкість до пошкоджень) досить висока. Так, пошкоджена частина фотографії не залишає можливості бачити частину, яка залишилася. Аналогові носії здатні зберігати інформацію на протязі тисяч років (єгипетські папіруси), сотень років (папір) та десятків років (магнітна стрічка, фото- і кіноплівка).

Цифрові носії більш чуттєві до пошкоджень, навіть втрата одного біту даних на магнітному або оптичному дисках може призвести до неможливості зрахувати файл, тобто до втрати великого обсягу даних. За експертними оцінками спеціалістів, при правильному зберіганні оптичні носії здатні зберігати інформацію сотню років, а магнітні – десятків років. Саме тому необхідно дотримуватися правила експлуатації і зберігання цифрових носіїв інформації.

4.2 Зберігання інформації в персональному комп'ютері

Комп'ютер - це прилад, який спеціально створений для роботи з інформацією. Інформація в комп'ютері зберігається в оперативній (внутрішній) пам'яті.

Пам'ять має бітову структуру, тобто складається із комірок, в які можна записати 0 або 1 (табл. 4.2).

Таблиця 4.2

Номер	Біти							
0	0	1	0	1	1	0	0	0
1	0	1	0	0	1	1	0	1 ...
2	1	0	1	1	0	1	1	0
3	0	0	1	0	1	1	0	0
...								

Біти об'єднали у групи по 8 – байти.

1 байт = 8 біт.

Кожний байт отримує порядковий номер – адресу.

Внутрішня пам'ять має наступні властивості: **дискретність** – пам'ять, яка складається з окремих комірок; **адресованість** – вісім розміщених підряд бітів пам'яті утворюють байт, всі байти пронумеровані, порядковий номер байта називається його адресою. Група з декількох байтів, які процесор може обробляти як єдине ціле, називають машинним словом (рис. 4.1). Довжина машинного слова може бути різною – 8, 16, 32 біти і т.д.

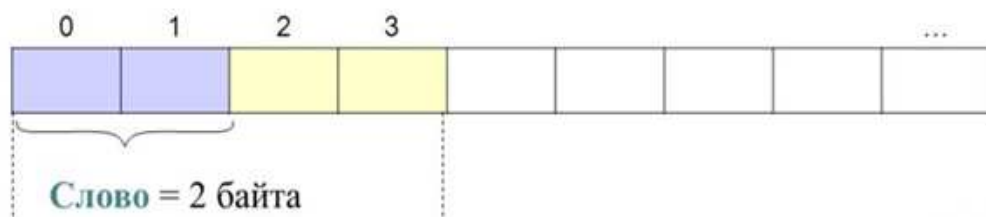


Рисунок 4.1

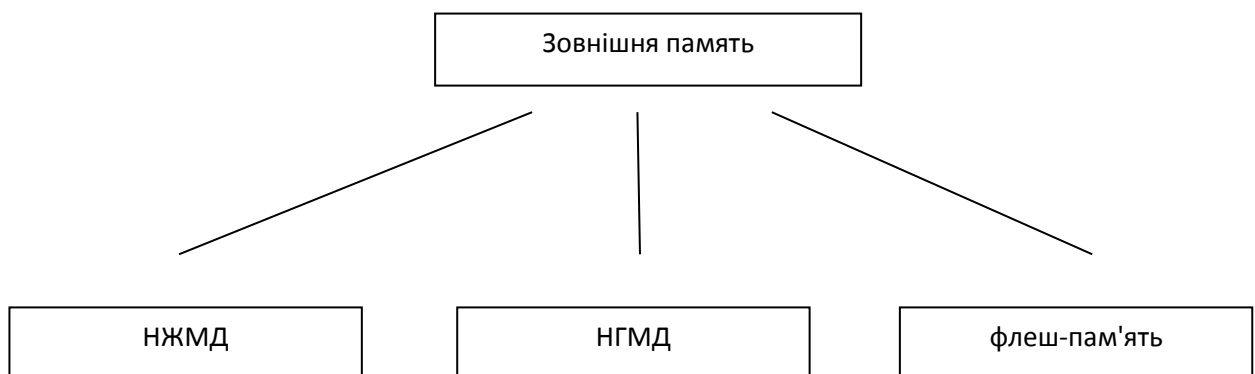





Рисунок 4.2

Зберігання інформації на зовнішніх носіях не вимагає постійного електроживлення, на відміну від внутрішньої пам'яті (рис. 4.2)

Таблиця 4.3 – Зовнішні носії інформації

	CD-ROM зберігає від 700 мбайт цифрової інформації
	Гнучкі диски (дискета) до 1,44 Мбайт
	Флеш-пам'ять від 512 Мбайт до 32 Гбайт

Оперативна пам'ять - пам'ять, яка призначена для тимчасового зберігання даних і команд, необхідних процесору для виконання ним операцій. Оперативна пам'ять виготовляється у вигляді модулів пам'яті (плоских пластин з електричними контактами, по бокам яких розміщуються великі інтегральні схеми пам'яті), які встановлюються на системній платі комп'ютера. Модулі пам'яті представляють собою пластини, на яких розміщуються великі інтегральні схеми (BIC) пам'яті. Модулі пам'яті можуть відрізнятися між собою за розміром і кількістю контактів, по швидкодії, по інформаційній ємності і т.д. У модулів оперативної пам'яті велика кількість показників (тип, вид, таймінги, частота), які суттєво впливають на роботу пам'яті.

При роботі пам'ять комп'ютера звертається до одного із двох типів «сховищ» інформації – оперативного запам'ятовуючого пристрою (ОЗП) та постійного запам'ятовуючого пристрою (ПЗП). ОЗП - це таке сховище інформації, яке повинне постійно відновлюватися, щоб у ньому зберігалася різна інформація, яка необхідна в даний момент для роботи комп'ютера. Вона автоматично очищається при відключенні комп'ютера від електроживлення.

Статична пам'ять комп'ютера - ПЗП (постійний запам'ятовуючий пристрій) - це сховище інформації, яке розраховане на незмінне та довготривале зберігання файлів, які повинні знаходитися у пам'яті комп'ютера, після того як комп'ютер буде відключений від електроживлення. Енергонезалежна пам'ять використовується для довготривалого зберігання

інформації і вимагає підключення електричної напруги (наприклад, батарейки). Карта енергонезалежної пам'яті представляє собою ВІС, яка розміщена у мініатюрному плоскому корпусі.

Для запису і зчитування інформації з карт пам'яті використовуються спеціальні пристрої, які не мають рухомих частин і тому забезпечують високу збереженість даних при використанні у мобільних пристроях (портативних комп'ютерах, цифрових камерах та інших).

Зовнішня (довготривала) пам'ять - це місце тривалого зберігання даних (програм, результатів розрахунків, текстів і т.д.), що не використовуються на даний момент в оперативній пам'яті комп'ютера. Для роботи із зовнішньою пам'яттю необхідна наявність накопичувача (дисководу - пристрою, що забезпечує запис і зчитування інформації) та пристрою зберігання - носія.

Пристрої прийнято поділяти на види і категорії у зв'язку з їх принципами функціонування, експлуатаційно-технічними, фізичними, програмними та іншими характеристиками. При вимиканні комп'ютера вся інформація з оперативної пам'яті стирається. Для довготривалого зберігання інформації використовується зовнішня пам'ять. Пристрій, який забезпечує записи і зчитування інформації, називається **накопичувачем** або **дисководом**, а зберігається інформація на носіях інформації. Інформація на носіях зберігається у цифровій формі, тобто у формі послідовностей нулів та одиниць. Якщо в якості носія інформації використовується дискета, всередині пластмасового корпусу якої розміщується гнучкий магнітний диск, то інформація на дискеті зберігається на концентричних доріжках, на яких чергуються намагнічені і ненамагнічені ділянки (намагнічена ділянка зберігає комп'ютерну «1», ненамагнічена - комп'ютерний «0»).

Для запису або зчитування інформації дискета вставляється у дисковод, який обертає диск всередині пластмасового корпусу дискети і магнітна голівка дисковода встановлюється на певну концентричну доріжку диску.

Інформаційна ємність дискети така, що на ній можна розмістити близько 600 сторінок тексту або кілька десятків зображень. З метою збереження інформації дискети необхідно оберігати від сильних магнітних полів.

Гнучкі магнітні диски. Знімні магнітні диски (дискети) вставляють в комп'ютер через спеціальну щілину системного блоку - дисковод. Насправді це не один диск, а група дисків, що мають магнітне покриття і обертаються з високою швидкістю. Основними параметрами гнучких дисків є: технологічний розмір (вимірюється в дюймах), щільність запису (вимірюється у кратних одиницях) і повна ємність.

Жорсткі магнітні диски або НЖМД, вінчестер - основне сховище інформації великих обсягів, основане на принципі магнітного запису, прихований всередині корпусу системного блоку. Є основним накопичувачем даних у більшості комп'ютерів. Інформація в НЖМД записується на жорсткі пластини, які покриті шаром феромагнітного матеріалу. Носій інформації суміщений з накопичувачем, приводами блоком електроніки і звичайно встановлений всередині системного блоку комп'ютера.

Зовнішні жорсткі диски - динамічні системи зберігання даних. Вони зручні при веденні бізнесу, надають свободу творчості, взаємодії в будь-який час, у будь-якому місці. Зовнішній жорсткий диск простий у використанні завдяки своїй портативності, підтримує високошвидкісний інтерфейс для швидкого передавання даних.

Оптичні дисководи і диски. Збірна назва для носіїв інформації, виконаних у вигляді дисків, читання з яких ведеться за допомогою оптичного випромінювання. Диски зазвичай плоскі, їх основа зроблена з полікарбонату, на який нанесений спеціальний шар для зберігання інформації. Для зчитування інформації використовується звичайно промінь лазера, який направляється на спеціальний шар і відбивається від нього.

Дисководи оптичних дисків зчитують інформацію разів в десять - п'ятнадцять швидше, ніж гнучкі, але все ж повільніше, ніж жорсткі. Зате вони

знімні і дуже об'ємні (на них поміщається приблизно в 500 разів більше інформації, ніж на гнучкий диск - до 700 Мб).

У направляючому лотку компакт-диск потрапляє всередину пристрою і закріплюється на шпинделі, що обертається. При обертанні диск освітлюється лазерною голівкою, промінь якого або відбивається від поверхні, або розсіюється. Лазерна голівка переміщається уздовж диску за допомогою позиціонера.

Відбитий промінь виробляє сигнал, що передається, в кінцевому підсумку, через оперативну пам'ять у центральний процесор для обробки.

Поверхня компакт-диску представляє собою одну спіральну доріжку, на якій розміщуються мікроскопічні впадини, що розсіюють лазерний промінь, який на них потрапляє.

Набір нулів і одиниць на диску розташовуються за правилами, які називаються форматом. Музичні компакт-диски - це просто один з форматів.

В залежності від того, де розташовується інформація - ближче до краю диска або ближче до центру, швидкість її зчитування змінюється. У маркуванні дисководів вказується максимальна швидкість читання.

Сучасні дисководи мають максимальну швидкість читання від 45 до 55Х.

У всіх лазерних дисководах від 1 до 12 - швидкісних використовувалася технологія CLV. При цьому дані з краю компакт-диску зчитуються з такою ж швидкістю, як і дані з внутрішніх доріжок. У більш швидкісних CD - ROM'ах використовується інша технологія, комбінація CLV і CAV, і дані зовнішніх доріжок зчитуються набагато швидше внутрішніх.

Використання тільки постійної лінійної швидкості призвело б до того, що при читанні внутрішніх областей диску він обертася б так швидко, що призвело б до руйнування або неправильної роботи дисководу. Таким чином, внутрішні області зчитуються в режимі CAV, а зовнішні - в режимі CLV.

Фабрично виготовлені компакт-диски штампуються на спеціальних пресах.

Перше покоління оптичних дисків: лазерний диск, компакт-диск, магнітооптичний диск.

Друге покоління оптичних дисків: DVD, MiniDisc, Digital Multilayer Disk, DataPlay, Fluorescent Multilayer Disc, GD-ROM (Compact Disk Read-Only Memory), Universal Media Disc.

Третє покоління оптичних дисків: Blu-rayDisc, HDDVD, Forward Versatile Disc, Ultra Density Optical, Professional Disc for DATA, Versatile Multilayer Disc.

Четверте покоління оптичних дисків: HolographicVersatileDisc, SuperRensDisc.

Лазерні дисководи і диски. Лазерні дисководи (CD-ROM і DVD-ROM) використовують оптичний принцип читання інформації. На лазерних CD-ROM (CD - CompactDisk, компакт-диск) і DVD-ROM (DVD - Digital Video Disk, цифровий відеодиск) дисках зберігається інформація, яка була записана на них у процесі виготовлення. Запис на них нової інформації неможлива, що відображено у другій частині їх назв: ROM (ReadOnlyMemory - тільки читання). Виробляються такі диски шляхом штампування і мають сріблястий колір. На дисках CD-RW і DVD-RW (RW - ReWntable, перезаписуваний), які мають «платиновий» відтінок, інформація може бути записана багаторазово.

У лазерних дисководах використовується оптичний принцип запису та зчитування інформації. Інформація на лазерному диску зберігається на одній спіралевидній ділянці (як на равлику), що містить ділянки, які чергуються, з поганою і гарною здатністю (не відображає - комп'ютерний «0», відображає - комп'ютерна «1»).

У процесі зчитування інформації з лазерних дисків лазер, який встановлений у дисководі, падає на поверхню диска, що обертається і відбивається. Так як поверхня лазерного диску має ділянки з різною відображаючою здатністю, відбитий промінь також змінює свою інтенсивність і перетворюється у цифровий комп'ютерний код.

З метою збереження інформації лазерні диски необхідно оберігати від забруднень і подряпин.

XDPicture (ExtremeDigital) - були створені компаніями FujiFilm і Olympus для заміни порядком застарілого формату SmartMedia. Застосовуються дані карти переважно в цифрових фотоапаратах цих компаній. Також останнім часом широкого розповсюдження набули USB флеш-накопичувачі («флешка», USB-драйв, USB-диск), практично витіснили дискети і CD.

Flash-пам'ять. Flash-пам'ять - це енергонезалежний тип пам'яті. Вона представляє собою мікросхему, яка вміщена у мініатюрний плоский корпус. Для зчитування або запису інформації карта пам'яті вставляється у спеціальні накопичувачі, вбудовані у мобільні пристрої або підключаються до комп'ютера через USB-порт. Карти flash-пам'яті не мають у своєму складі рухомих частин, що забезпечує високу збереженість даних при їх використанні у мобільних пристроях (портативних комп'ютерах, цифрових камерах та ін.). Їх існує велика кількість: SD, MMC, CompactFlashType I і II, MemoryStick, MemoryStickDuo, TransFlash, miniSD, microSD, RS-MMC, SmartMedia, MiniDisk та ін.

CompactFlash - флеш-пам'ять. Всього існує два типи карт CompactFlash: CF Type I, CF Type II, причому відрізняються вони тільки товщиною корпусу.

SD (SecureDigital) - також був створений зусиллями компаній SanDisk, Panasonic і Toshiba. У цих картах використовуються криптограми (шифрування даних), що забезпечує захист даних від несанкціонованого копіювання або перезапису.

MMC (MultiMediaCard) - результат роботи компаній SanDisk і Siemens. У кожній MMC є власний контролер пам'яті. При цьому товщина мультимедійних карт майже на третину менше, ніж у «шпигунського» брата, що дозволяє використовувати MMC-накопичувачі у різних мініатюрних пристроях.

RS-MMC (ReducedSize MMC) - також відомі як MMCmobile. Вони відрізняються від MMC лише зменшеними розмірами і використовуються в основному у мобільних телефонах.

Memory Stick Duo - є еволюцією самих Memory Stick. Зменшилися розміри та енергоспоживання карт, але разом з тим зменшилася і максимальна ємність. В іншому повністю аналогічна звичайній MS.

SmartMedia - стандарт, розроблений Toshiba. Особливостями даного стандарту можна вважати дуже низьке енергоспоживання і відсутність власного контролера, швидкість роботи вкрай низька і максимальний обсяг пам'яті становить всього-на-всього 256 Мб.

XDPicture (ExtremeDigital) - створені компаніями FujiFilm і Olympus для заміни порядком застарілого формату SmartMedia. Застосовуються дані карти переважно у цифрових фотоапаратах цих компаній.

Також останнім часом широкого поширення набули USB флеш-накопичувачі («флешка», USB-драйв, USB-диск).

4.3 Оперативна пам'ять

У комп'ютері теж є декілька засобів для зберігання інформації. Найшвидший спосіб запам'ятати дані - це записати їх в електронні мікросхеми. Така пам'ять називається **оперативною пам'яттю**. Це швидкий запам'ятовуючий пристрій не дуже великого об'єму, який безпосередньо пов'язаний з процесором та призначений для запису, зчитування та зберігання даних та програм, що виконуються, програм, які обробляються. Оперативна пам'ять використовується тільки для тимчасового зберігання даних та програм під час роботи, енергозалежна, після вимикання її все, що знаходилося в ОЗП, пропадає.

Оперативна пам'ять складається з комірок (рис. 4.3). У кожній комірці може зберігатися один байт даних. У кожної комірки є свої **адреси**. Можна вважати, що це як би номер комірки, тому такі комірки ще називають **адресними комірками**. Коли комп'ютер відправляє дані на зберігання в оперативну пам'ять, він запам'ятовує адреси, в які ці дані поміщені. Звертаючись до адресної комірки, комп'ютер знаходить в ній байт даних. Дані в

оперативній пам'яті зберігаються байтами. Кількість байтів, які можна зберегти в оперативній пам'яті, залежить від її обсягу. Об'єм оперативної пам'яті вимірюють кілобайтами (кбайт) або Мегабайтами (Мбайт).

	0	1	2	255
0	00000	00001	00002	00255
1	00256	00257	00258	00511
2	00512	00513	00514	00767
.
.
.
.
255	65280	65281	65282	65535

Рисунок 4.3

Двома байтами (від 0 до 255) можна записати адресу для 65536 комірок пам'яті (0 до 65535). Для більшої кількості комірок адреса повинна мати більше байтів. Умовно вважають, що кілобайт дорівнює тисячі байтів. Насправді 1 кбайт дорівнює 2^{10} , тобто 1024 байтам. Так само вважають, що один мегабайт дорівнює тисячі кілобайтів або мільйону байтів хоча більш точно 1 Мбайт = 1024 кбайт = 1 048 576 байт (2^{20}).

Регенерація оперативної пам'яті

Адресна комірка оперативної пам'яті зберігає один байт, а оскільки байт складається з восьми бітів, то в ній є вісім **бітових комірок**. Кожна бітова комірка мікросхеми оперативної пам'яті зберігає електричний заряд. Заряди не можуть зберігатися у комірках довго. Всього за декілька десятих долей секунди заряд у комірці зменшується настільки, що дані втрачаються. Для того, щоб не забути інформацію, комп'ютер її регулярно повторює. Десятки разів за секунду він перевіряє, що міститься у комірках пам'яті і «підзаряджує» кожен комірку (як би повторює запис). Це називається **регенерацією оперативної пам'яті**. Регенерація пам'яті відбувається дуже швидко. Ми не помічаємо, як кожен секунду кілька разів оновлюються мегабайти пам'яті, але варто тільки

на [мить](#) вимкнути живлення комп'ютера, як [регенерація](#) припиниться. Навіть короточасне зникнення напруги в [мережі](#) призводить до стирання оперативної пам'яті і «скидання» комп'ютера.

Пам'ять на магнітних дисках

Мікросхеми оперативної пам'яті запам'ятовують і видають дані дуже швидко, тому вони використовуються для обробки інформації, але для тривалого зберігання даних вони не пригодні - тут потрібні інші способи. Коли людині треба щось міцно запам'ятати, вона використовує записну книжку. Комп'ютер теж має "записні книжки" - це [магнітні](#) диски. [Магнітні](#) диски бувають двох типів - гнучкі і жорсткі.

Гнучкі диски (дискети) мають не дуже велику ємність і працюють порівняно повільно, але їх можна переносити з одного комп'ютера на інший.

У кожному комп'ютері є один або кілька дисків (їх називають вінчестерами). **Жорсткий диск** – це магнітний диск, який встановлюється у системному блоці комп'ютера (рис. 4.4). Зовнішньо цей диск представляє собою герметичну металеву коробку, всередині якої розміщений сам диск, магнітні головки читання-запису, механізми обертання диску та переміщення головок. Жорсткий диск складається з декількох дисків, що нанизані на загальну ось. Запис інформації здійснюється на обидві сторони кожного диску. Жорсткі диски відрізняються один від одного перед усім своєю ємністю. Сучасні жорсткі диски мають ємність від одного до декількох сотень гігабайт (Гб). Жорсткі диски мають велику ємність, але вони розташовуються всередині системного блоку і їх не можна переносити. Диск обертається з величезною швидкістю, а над магнітною [поверхнею](#) розміщується на повітряній подушці [магнітна](#) голівка, яка записує і зчитує біти і байти даних. Корпус жорсткого диска закритий кожухом, знімати який не можна, інакше мікрочастинки пилу, які туди потраплять, з часом виведуть диск з ладу.



Рисунок 4.4 – Жорсткий диск

Жорсткі магнітні диски представляють собою декілька тонких металевих дисків, які дуже швидко обертаються на одній осі і поміщені в металевий корпус. Магнітне покриття дисків містить сотні тисяч концентричних доріжок, на яких зберігається інформація.

Інформаційна ємність жорстких дисків дуже велика - на одному жорсткому диску може бути розміщена ціла бібліотека, що складається з десятків книг. З метою збереження інформації жорсткі диски необхідно оберігати від ударів.

Жорсткі диски, на відміну від гнучких, як правило, незнімні. Жорсткий магнітний диск зі знятим кожухом і накопичувач на жорстких магнітних дисках (НЖДМ, HDD-hard disk drive).

На жорстких магнітних дисках феромагнітний порошок нанесений не на тонку пластикову основу, а на алюмінієвую або (в останніх моделях) на скляну. На одному шпинделі, який обертається, кріпиться цілий пакет дисків (до трьох), до кожної сторони яких підходить магнітна голівка. Таким чином, при фіксованому положенні позиціонера всі голівки описують кілька концентричних кіл, які складають один циліндр.

Позиціонер представляє собою кроковий електродвигун, що має кілька сотень (або навіть тисяч) фіксованих положень. Кількість таких положень і визначає кількість на НЖМД.

На жорстких дисках може зберігатися велика кількість інформації, зокрема фотографії, відео, музика, документи, програми тощо. На жорсткому диску комп'ютера інформація зберігається навіть тоді, коли комп'ютер

вимкнено. Існують також зовнішні жорсткі диски. Найпростіший спосіб збільшення обсягу вільного місця на ПК — під'єднання зовнішнього жорсткого диска. Він може доповнювати обсяг внутрішнього жорсткого диска комп'ютера, особливо якщо внутрішній диск функціонує повільніше через нехватку місця. На зовнішньому жорсткому диску можна зберігати резервну копію важливих файлів.

Існують також зовнішні жорсткі диски. Найпростіший спосіб збільшення обсягу вільного місця на ПК — під'єднання зовнішнього жорсткого диска. Він може доповнювати обсяг внутрішнього жорсткого диска комп'ютера, особливо якщо внутрішній диск функціонує повільніше через нехватку місця. На зовнішньому жорсткому диску можна зберігати резервну копію важливих файлів. Більшість таких жорстких дисків під'єднується до USB-порту.

Характеристики накопичувачів на жорстких дисках:

- ємність;
- швидкодія;
- надійність;
- вартість.

Ємність. Після того, як користувач повністю заповнює весь вільний простір поточного жорсткого диска, він починає замислюватися про те, який обсяг пам'яті буде достатнім. Ймовірність того, що наявного простору виявиться занадто багато, досить незначна, тому необхідно придбати найбільший жорсткий диск. Сучасні системи використовуються для зберігання об'ємних файлів різних форматів, до числа яких відносяться цифрові фотографії, музичні записи й відеофрагменти, новітні операційні системи, додатки та комп'ютерні ігри. У сучасних системах нестача вільного місця призводить до виникнення різних проблем, які пов'язані головним чином з тим, що операційна система Windows і прикладні програми використовують великий обсяг дискового простору для віртуальної пам'яті й зберігання тимчасових

файлів. Вихід Windows за межі ємності жорсткого диска практично завжди призводить до нестійкої роботи системи, збоєм та втраті даних.

Швидкодія. Швидкодія накопичувача можна оцінити по двом параметрам: швидкості передавання даних (data transfer rate); середньостатистичному часу пошуку (average seek time).

Швидкість передавання даних. Швидкість передавання даних показує, наскільки швидко виконується запис або зчитування даних, що знаходяться на жорсткому диску. Швидкість передавання даних носія може бути виражена у вигляді повної швидкості (максимальної або мінімальної), максимальної або мінімальної фактичної швидкості, а також у вигляді середньої фактичної швидкості. Середня швидкість передавання даних вважається більш важливою характеристикою, ніж швидкість передавання даних інтерфейсу. Це пов'язане з тим, що середня швидкість представляє собою дійсну швидкість безпосереднього зчитування даних з поверхні жорсткого диска. При цьому максимальна швидкість є, скоріше, очікуваною постійною швидкістю передавання даних.

Середній час позиціонування. Середній час позиціонування, що вимірюється в мілісекундах (мс), - це час, який необхідний для переміщення голівки від одного циліндру до іншого на довільну відстань. Середній час позиціонування залежить безпосередньо від конструкції жорсткого диска. Величина середнього часу позиціонування відображає можливості механізму приводу головки.

Час очікування. Часом очікування називається середній час (у мілісекундах), який необхідний для переміщення голівки до зазначеного сектора після досягнення голівкою певної доріжки. У середньому ця величина дорівнює половині часу, що вимагається для одного оберту жорсткого диска. При збільшенні частоти обертання диска вдвічі час очікування зменшиться наполовину. Час очікування є одним із факторів, що визначають швидкість читання й запису накопичувача. Зменшення часу очікування призводить до зменшення часу доступу до даних або файлів.

Середній час доступу. Середнім часом доступу до даних (у мілісекундах) називається сума середнього часу позиціонування й часу очікування. Величина середнього часу доступу (середній час позиціонування разом з часом очікування) представляє собою середній час, який необхідна накопичувачу для звертання до довільно розташованого сектора.

Надійність Це середньостатистичний час між збоями {Mean Time Between Failures - MTBF), що коливається від 20 до 500 тис. годин і більше.

Щоб дані можна було не тільки записати на жорсткий диск, а потім ще й прочитати, треба точно знати, що і куди було записано. У всіх даних повинен бути **адреса**. У кожної книги в бібліотеці є свій зал, стелаж, полиця та [інвентарний](#) номер - це як би її адреса. По такій адресі книгу можна знайти. Всі дані, які записуються на жорсткий диск, теж повинні мати адресу, інакше їх не знайти. Якщо запам'ятовувати окремо кожну адресу, в якій були записані байти даних, то зберігати ці адреси стане важче, ніж самі дані. Інформація зберігається не байтами, а файлами. **Файл** - найменша одиниця зберігання даних. Кожен [файл](#) на диску має свою адресу. Якщо нам потрібна якась інформація, комп'ютер знаходить на диску потрібний [файл](#), а потім байт за байтом зчитує з нього дані в оперативну пам'ять, поки не дійде до кінця файлу.

Щоб у кожного файла на диску була своя адреса, диск розбивають на **доріжки**, а доріжки, у свою чергу, розбивають на **сектори**. Розмір кожного сектора [стандартний](#) і дорівнює 512 байтам. Розбиття диска на доріжки та сектори називається **форматуванням диску**. Його виконують службові програми. Форматування диску чимось схоже на розлінування зошиту. [Як](#) і для зошиту, форматування диску потрібно виконати тільки один раз.

Найперша доріжка магнітного диску (нульова) вважається службовою - там зберігається службова інформація. Наприклад, на цій доріжці зберігається таблиця розміщення файлів (FAT-таблиця). У цій [таблиці](#) комп'ютер запам'ятовує адреси записаних файлів. Коли нам потрібний якийсь файл, комп'ютер по його імені знаходить у цій таблиці номер доріжки і номер

сектору, після чого магнітна головка переводиться у потрібне положення, файл зчитується і направляється в оперативну пам'ять для обробки.

Якщо таблиця розміщення файлів буде пошкоджена, то інформація, яка зберігається на диску, може бути втрачена. Насправді вона там залишається, але до неї не можна звернутися. Тому таблиця розміщення файлів для надійності дублюється. У неї є копія, і при будь-яких [пошкодженнях](#) комп'ютер сам відновлює цю таблицю. Завдяки цьому з комп'ютером можна працювати роками і не втрачати дані.

Розміщення файлів на жорсткому диску

У кожного файла є своя **адреса**. Ця адреса записана у таблиці розміщення файлів двухбайтовим числом, тобто, на запис цієї адреси надано 16 бітів (тому таблицю розміщення файлів ще називають **FAT 16**). За допомогою 16 бітів можна виразити 2^{16} (65536) різних значень. Це означає, що файлів на жорсткому диску не може бути більше, ніж 65536 різних адрес (самих файлів не може бути більше 65536).

Сучасні [жорсткі диски](#) мають дуже великі обсяги, і їм не вистачає такої кількості адрес (табл. 4.4). Якщо, наприклад, розмір диска 2 Гбайт (два мільярди байтів), то на кожен адресу припадає $2 \text{ Гбайт} / 65536 = 32 \text{ кбайт}$.

Кожний файл повинен мати власну унікальну адресу. Мінімальний розмір адресного простору називається **кластером**. Для жорстких дисків, що мають розмір 2 Гбайт, кластер дорівнює 32 кбайти. Якщо диск менший, то й кластер у нього теж менший. Для дисків 1 Гбайт кластер дорівнює 16 кбайт.

У сучасних дисків кластер набагато більше сектору, який дорівнює 0,5 кбайт. В одному кластері можуть міститися десятки секторів, і, яким би маленьким не був файл, він все одно займе цілий кластер, і всі сектори, які не використовуються, у ньому просто пропадуть.

Таблиця 4.4 - Зв'язок між розміром жорсткого диска і розміром кластеру

Об'єм диску	Розмір кластера
Менше 32 Мбайт	512 байт
32 Мбайт...64 Мбайт	1 кбайт
64 Мбайт...128 Мбайт	2 кбайт
128 Мбайт...256 Мбайт	4 кбайт
256 Мбайт...512 Мбайт	8 кбайт
512 Мбайт...1 Гбайт	16 кбайт
1 Гбайт...2 Гбайт	32 кбайт

Найближчим часом комп'ютери перейдуть на нову систему запису адреси файлу на жорсткому диску, яка називається **FAT 32** (табл. 4.5). За назвою неважко здогадатися, що в цій системі адреса записується не двома байтами, а чотирма (32 біти). Тоді адрес стане набагато більше, а розміри окремих [кластерів](#) - менше. Нераціональні втрати набагато зменшаться.

Таблиця 4.5 - Розміри кластерів для FAT 32

Об'єм диску	Розмір кластера
513 Мбайт...8 Гбайт	4 кбайт
8 Гбайт...16 Гбайт	8 кбайт
16 Гбайт...32 Гбайт	16 кбайт
Більше 32 Гбайт	32 кбайт

Диски фізичні і логічні

Всі, хто працюють з комп'ютером, звикли цінувати робоче місце на жорсткому диску. Чим більше диск, тим більше корисних і цікавих програм і даних можна розмістити на ньому. З іншого боку, виходить, що чим більше жорсткий диск, тим більше місця на ньому пропадає марно через недосконалу систему адресації файлів.

Для боротьби з нераціональними втратами жорсткий диск розбивають на декілька розділів. Для цього є спеціальні програми. Наприклад, на жорсткому диску, що має розмір 2 Гбайт, створюють чотири розділи по 0,5 Гбайт. Кожний [такий](#) розділ можна розглядати як один окремий **логічний жорсткий диск**. Звичайний жорсткий диск - це фізичний пристрій. Його можна [встановити](#) або видалити. Логічний жорсткий диск фізично не існує. Це просто один з розділів фізичного диску. Працюючи з комп'ютером, користувач не помічає різниці між фізичними і логічними дисками. Кожний логічний диск має власну таблицю розміщення файлів, тому на ньому діє своя система адресації. В результаті втрати із-за розмірів кластерів стають меншими.

Імена дисків

Кожний диск, що присутній на комп'ютері, має унікальне ім'я. Неважливо, що це за диск: фізичний або логічний, - у нього обов'язково має бути ім'я. Ім'я диску складається з однієї літери англійського алфавіту і двокрапки, наприклад, A:, F:. Коли на комп'ютері встановлюється новий жорсткий диск, він отримує літеру наступну за останньою використаною буквою. Те ж [саме](#) відбувається і при створенні нового логічного диску на вже встановленому фізичному диску. Літерою A: загальноприйнято позначати дисковод для гнучких дисків. Літерою C: позначається перший жорсткий диск. Наступний диск отримує літеру D:, потім E: і т.д. Літера V: зарезервована на той [випадок](#), що в комп'ютері може бути не один, а два дисководи гнучких дисків.

Адреса файлу. Поняття про каталог

Для зручності [роботи](#) з файлами на диску створюються *каталоги*. Якщо всередині одного каталогу знаходиться інший, то їх імена відокремлюються один від одного зворотного косою рисою (\). Кожний сам створює собі на диску такі каталоги, які йому зручні.

Наприклад, на диску C: можна створити наступні каталоги:

C:\Проекти;

C:\Статті;

C:\Архіви.

Якщо каталог знаходиться всередині іншого каталогу, він називається **вкладеним**.

У різних операційних системах різні правила запису імен файлів. Адресу файлу ще називають **шляхом доступу до файлу** або **шляхом пошуку файлу**. Знаючи таку адресу (шлях), неважко знайти будь-який файл з наявних на комп'ютері. Потрібно тільки знати, де що знаходиться. Але це вже завдання не для комп'ютера, а для користувача.

Пристрої читання CD і DWD-дисків

Майже всі комп'ютери сьогодні обладнано пристроями читання CD і DWD-дисків (приводами), які зазвичай розташовано на передній панелі системного блоку. За допомогою лазерів пристрої зчитують (отримують) дані з диска. Ці пристрої можуть також записувати інформацію на диски.

Завдяки наявності в комп'ютері пристрою читання оптичних дисків, можна переглядати фільми або прослуховувати музичні DWD-диски.

Флеш-пам'ять або просто **флешка** — це маленький портативний пристрій у вигляді брелка, який підключається до USB-порту комп'ютера. Подібно до жорсткого диска, флеш-пам'ять використовується для зберігання інформації, проте зазвичай її обсяг значно менший порівняно з більшістю жорстких дисків. Флеш-пам'ять може мати різний розмір, форму та зберігати гігабайти даних. Такі пристрої зручно носити із собою, що дає змогу використовувати їх для перенесення інформації з одного комп'ютера на інший. Флеш-пам'ять також називають флешкою, флеш-брелком, флеш-накопичувачем.

У цифрових фотокамерах, телефонах інформація зберігається на картах пам'яті (флеш-пам'яті). Інформацію на картах пам'яті можна стирати та використовувати їх знову. Ноутбуки та деякі комп'ютери обладнано вбудованими пристроями читання карт пам'яті.

4.4 Зберігання інформації в Інтернеті

Інтернет - це об'єднання комп'ютерів по всьому світу в єдину інформаційну мережу. По-іншому Інтернет називають світовою комп'ютерною мережею. Для з'єднання комп'ютерів використовують звичайні телефонні лінії та модем. Модем перетворює інформацію у вигляд, придатного для передавання по телефону. Таким чином, інформація, що зберігається по всьому світі, стає доступною кожному, хто має комп'ютер, телефон і модем.

Телефонний зв'язок не є єдиним способом з'єднання комп'ютерів. Набагато швидше інформація передається по оптичних кабелях і за допомогою радіозв'язку. Ці канали поступово витісняють в Інтернет телефонні з'єднання.

В Інтернеті можна знайти відповідь практично на будь-яке питання. Прочитати свіжу газету, заглянути в бібліотеку, замовити квитки на літак, купити товари, завести друзів по листуванню. Програми і дані в комп'ютері зберігаються на жорсткому диску у вигляді файлів.

Файл - це певна кількість інформації, що має ім'я і зберігається у довготривалій (зовнішній) пам'яті.

Ім'я файлу - послідовність символів, що дозволяє користувачу орієнтуватися у файловій системі. Файл складається з двох частин, які розділені крапкою: ім'я файлу і розширення, що визначає його тип. Власне ім'я файлу може містити від 1 до 255 символів. Крім латинського допускається застосування російського алфавіту.

Розширення - це поєднання букв і чисел довжиною від одного до трьох символів, який доповнює саме ім'я, але частіше вказує на формат і тип даних, які зберігаються у файлі. Від імені файлу воно відокремлюється крапкою і є його необов'язковою частиною. Розширення використовуються для ідентифікації типу (формату) файлу. За їх допомогою користувач і програмне забезпечення комп'ютера може визначити тип даних, що зберігаються у файлі.

Розширення прийнято вказувати у вигляді *.rar, тобто перед символами розширення додають зірочку і точку, де зірочка символізує будь-яке ім'я файлу.

Розширення може вказувати не тільки на тип інформації, яка зберігається у файлі (зображення, медіа файл, текстовий файл), але і на спосіб кодування цієї інформації. Наприклад, *.gif, *.jpg, *.bmp, *.raw, *.png та ін. - це розширення файлів зображень, але способи кодування зображення у таких файлах різні, і не кожна програма, що відкриває один тип, зможе відкрити інший.

Існують файли, які не мають розширення, звичайно це системні файли. Файл відкривається тією програмою, у якій був створений, або універсальною програмою.

Приклади розширень файлів різних типів

* .doc, * .xdoc, * .rtf, * .txt, * .pdf - текстові документи (вміст таких файлів текст і відкриваються вони у програмі для роботи з текстом - Лист.doc, Каталог.xls, текст.txt);

* .jpg, * .gif, * .jpeg, * .bmp, * .raw, * .png, * .emf, * .ico, * .tif, * .tiff, * .jp2, * .pcx, * .tga, * .wbmp - графічне зображення (фотографії і картинки - Рисунок.gif, Природа.tif, Фото.jpg, Рисунок.bmp);

* .html, * .htm, * .xhtml - інтернет-сторінка (Книга.htm, Windows.hlp, Книга.html);

* .exe, * - файл, який виконується (містять програми готові до виконання - ACDSee9.exe, Command);

* .rar, * .zip, * .7z - заархівований файл (Реферат.zip, Реферат.rar);

* .avi, * .mpeg, * .mpg, * .mkv, * .3gp, * .mp4, * .wmf - розширення відеофайлів (вміст таких файлів відео і відкриваються вони у програвачах - Кліп.avi, Фільм.mpeg);

* .mp3, * .mp2, * .wave, * .wav, * .amr, * .wma, * .aac, * .ogg, * .midi, * .mid, * .kar - розширення аудіофайлів (музичний файл і відкривати його необхідно у програмі-програвачі - Пісня.mp3).

Файли із схожим вмістом або однаковим призначенням можна складати у папки (каталоги). Каталоги (папки), у свою чергу також можна об'єднувати у каталоги. Таким чином, на диску формується впорядкована структура файлів і

каталогів, яка дозволяє швидко знаходити потрібну інформацію. Така структура називається файловою системою.

Файлова система - частина операційної системи для зберігання файлів і організації каталогів, яка необхідна для упорядкованого розміщення даних і програм у зовнішній пам'яті. Файлова структура може бути багаторівневою і однорівневою.

Робота в [Інтернеті](#) іноді вимагає збереження потрібних даних та їх друку. Найпростіший спосіб копіювання даних з веб-сторінки - це копіювання тексту та вставка його у документи програм Office. Даний прийом виконується [стандартним](#) копіюванням через буфер обміну.

Для копіювання тексту з веб-сторінки слід виділити потрібний фрагмент, помістити у буфер обміну даних командою «Правка-Копіювати» (або поєднанням клавіш CTRL+C). Потім відкрити програму, в якій цей [текст](#) слід помістити і в потрібному місці дати команду вставки: «Правка-Вставити» (або поєднання клавіш CTRL+V). Для швидкої збірки текстового документа дуже зручні [стандартні програми Windows-блокнот](#) і WordPad. При цьому слід враховувати, що графічні елементи, стильове оформлення пропадають. Якщо є потреба збереження всіх даних веб-сторінки, рекомендується використовувати Microsoft Word. Дана програма зберігає стильове оформлення, графічні елементи і [гіперпосилання](#), але при цьому обсяг документа виходить значним, а зберегти документ без спотворень виходить тільки у форматі Word.

Іноді перенесення вмісту веб-сторінки в іншу програму неприйнятно для користувача, тоді допомагає [функція](#) збереження сторінки цілком або її частини у форматі веб-документа. Для збереження даємо команду «[Файл](#)-Зберегти як». У діалоговому вікні вказується місце збереження, ім'я файлу і формат, в якому буде зберігатися документ. Користувачеві пропонується 4 варіанти:

1. Формат Веб-сторінка, повністю - сторінка зберігається повністю разом із шрифтами, які використовуються, графікою, стилями та гіперпосиланнями. Відкривши [такий](#) документ, можна побачити сторінку у початковому вигляді у вікні браузера. Вихідна структура посилань не

зберігається, при збереженні створюється окрема папка із зображеннями і додатковими файлами, необхідними для повноцінного перегляду веб-сторінки;

2. Формат Веб-сторінка, тільки [HTML](#) зберігає сторінку без зображень, структура посилань початкової сторінки зберігається;

3. Формат Текстові файли - зберігає сторінку у вигляді текстового файлу, структура посилань порушується.

4. Формат Всі файли - використовується для збереження в будь - якому іншому форматі.

Багато веб-сторінок використовують фрейми - окремі блоки на сторінці. Дані у фреймах можуть формуватися окремо від вмісту сайту. Тому виникає необхідність відокремлено зберігати інформацію у фреймах. Для збереження окремого фрейму на веб-сторінці варто помістити курсор миші в даний фрейм і дати команду: «Файл-Зберегти фрейм як». Збереження відбувається також як збереження веб-сторінки.

Сучасний Інтернет насичений різноманітною інформацією: статтями, книгами, малюнками, [фотографіями](#), [анімацією](#) та ін. Використання широкосмугового доступу дозволяє користувачам безперешкодно копіювати будь-яку інформацію. Єдиним обмеженням тут є дотримання авторських прав власників.

Існують певні правила безпечного зберігання даних:

1. **Встановити паролі.** Серед усього переліку заходів безпеки, яких повинен дотримуватися користувач, перше місце займає його особиста організованість і відповідальне ставлення до важливої інформації, яку він зберігає у ПК. Найпростіший захід збереження конфіденційності даних — встановлення пароля для входу в систему комп'ютера. Звичайно, це не гарантує абсолютної безпеки, але, як мінімум, непередбачуваний або випадковий «недоброзичливець» не зможе просто так проникнути у Ваш ПК, навіть якщо він матиме на це багато часу. Аналогічне блокування можна створити й для файлів, якщо їх попередньо архівувати.

За допомогою спеціальних алгоритмів архіватори видаляють з файлу всю надмірну інформацію при стисненні. При зворотній операції розпаковування вони відновлюють інформацію у первісному вигляді. І стиснення, і відновлення інформації відбувається без втрат. При цьому можна задати пароль на розпаковування з архіву. Деякі архіватори надають можливість шифрувати не лише дані файлів, але й інші важливі області архіву: назви файлів, їхні розміри, атрибути, коментарі й інші блоки. Таке архівування з паролем часто використовують для захисту конфіденційної інформації в електронному листуванні: заархівований файл з паролем прикріплюють як вкладення до майже порожнього листа.

Щоб створити надійний пароль, використовують генератор паролів. Він створює випадковим чином паролі високого рівня надійності. Їх складно підібрати через використання в них великих і малих літер, чисел, знаків пунктуації та інших знаків. Створені таким чином паролі ніде не будуть збережені, якщо їх не зробити самостійно. Потрібно пам'ятати, що який би складний пароль на архівовані файли не був встановлений, ці файли можна просто видалити і, таким чином, втратити важливі дані;

2. Розділити жорсткий диск на кілька розділів (логічних дисків): в одному з них зберігати програмне забезпечення й системні дані, в інших — решту інформації;

3. Створити кілька облікових записів користувачів: лише один з них повинен мати права адміністратора, тобто права встановлювати і видаляти програмне забезпечення;

4. При використанні ОС Windows встановити Linux. Windows «не бачить» файлової системи ext4, яку використовує Linux. Тому при ураженні вірусом програмного забезпечення на платформі Windows, завантаживши Linux, можна «витягнути» з ПК усі напрацювання до перевстановлення Windows. Лікування деяких вірусів, що уражають флеш-карти при роботі на ПК під управлінням Windows, зводиться до простого вилучення файлів із флеш-карти при роботі Linux.

5. Використовувати антивірус. Однією з головних причин втрати даних (друге місце після фактів ненавмисного видалення) є наслідки дій шкідливого програмного забезпечення. Обов'язково потрібно встановити антивірус. Спеціалізоване сучасне ПЗ, у тому числі безкоштовне, дозволить мінімізувати загрозу, вчасно її виявити, а в разі проникнення — блокувати, вилікувати або видалити. Основні напрямки захисту від комп'ютерних вірусів є такі: запобігання надходженню вірусів; запобігання вірусній атаці, якщо вірус потрапив у комп'ютер; запобігання руйнівним наслідкам, якщо атака відбулася.

6. Регулярно створювати резервні копії. На жаль, інформацію неможливо абсолютно надійно зберігати в комп'ютері. Відмова апаратної частини (жорсткий диск), вірусна атака чи неакуратність самого користувача (випадкове видалення інформації) можуть призвести до втрати важливих даних. Щоб убезпечити важливу інформацію, необхідно робити резервне копіювання даних. **Резервне копіювання даних** — це створення копій важливої інформації, що є на ПК, для збереження її в інших сховищах даних (флеш-накопичувач, жорсткий диск, DVD-диск, хмарний сервіс тощо). Резервне копіювання файлів дозволить захистити дані у випадку, якщо станеться вихід з ладу основного носія інформації (наприклад, жорсткого диску комп'ютера), або вірусної атаки. Розрізняють такі типи резервних копій:

- Резервна копія операційної системи. Дуже корисна річ, якою часто нехтують навіть досвідчені користувачі. Потрібно встановити операційну систему, драйвери і необхідні програми. Потім зробити резервну копію налагодженої операційної системи і, у разі потреби (вірусна атака, або просто захаращення системи), відновити резервну копію. Це займає набагато менше часу, ніж нова установка і налаштування операційної системи. Як правило, для резервного копіювання операційної системи використовують спеціальні програми або засоби;

- Резервна копія логічного диску (розділу);
- Резервна копія окремих файлів — найпоширеніший спосіб резервного копіювання.

Резервне копіювання та відновлення даних є різновидом операцій збереження даних із своїми особливостями:

- при збереженні даних ми найчастіше маємо справу з одним або декількома файлами. При резервному копіюванні зазвичай об'єктом копіювання є набір великої кількості файлів або диски. Інакше кажучи, це комплексна дія;
- резервне копіювання завжди проводять на інші носії — на DVD-диск, флеш-накопичувач, мережні ресурси інформації, відмінні від тих, з яких копіюють;
- при резервному копіюванні значення збережених даних для користувача комп'ютера більше, ніж у випадку окремого файлу або файлів. Тому, для таких випадків використовують спеціальні засоби - програми для резервного копіювання даних.

4.5 Основні системи зберігання даних та їх особливості

Мережа зберігання даних Storage Area Network, SAN

З повсякденним розвитком мережних комп'ютерних систем і глобальних корпоративних рішень світ почав вимагати технологій, які б дали поштовх до розвитку корпоративних систем зберігання інформації (сторедж-систем). І ось, одна єдина технологія приносить у світову скарбницю досягнення в області сторедж швидкодію, можливості масштабування і виняткові переваги загальної вартості володіння. Обставини, які сформувалися з появою стандарту FC-AL (Fibre Channel - Arbitrated Loop) і SAN (Storage Area Network), яка розвивається на його основі, обіцяють революцію у дата-орієнтованих технологіях комп'ютингу.

Мережа зберігання даних (SAN) представляє собою архітектурне рішення для підключення зовнішніх пристроїв зберігання даних, таких як дискові масиви, стрічкові бібліотеки, оптичні накопичувачі до серверів таким чином, щоб операційна система розпізнала підключені ресурси, як локальні. Побудова мережі SAN вирішує проблеми зменшення сукупної вартості

володіння системою зберігання даних, а також надає інструменти для організації надійного зберігання інформації. SAN - мережа, основним завданням якої є передавання даних між комп'ютерними системами і пристроями зберігання даних, а також між самими сторедж-системами. SAN складається з комунікаційної інфраструктури, яка забезпечує фізичний зв'язок, а також відповідає за рівень управління (management layer), який об'єднує зв'язки, сторедж і комп'ютерні системи, здійснюючи передавання даних безпечно і надійно.

У найпростішому випадку SAN складається із **системи зберігання даних** (СЗД), комутаторів і серверів, які об'єднані оптичними каналами зв'язку. Крім безпосередньо дискових СЗД у SAN можна підключити дискові бібліотеки, стрічкові бібліотеки (стрімери), пристрої для зберігання даних на оптичних дисках (CD/DVD та інші).

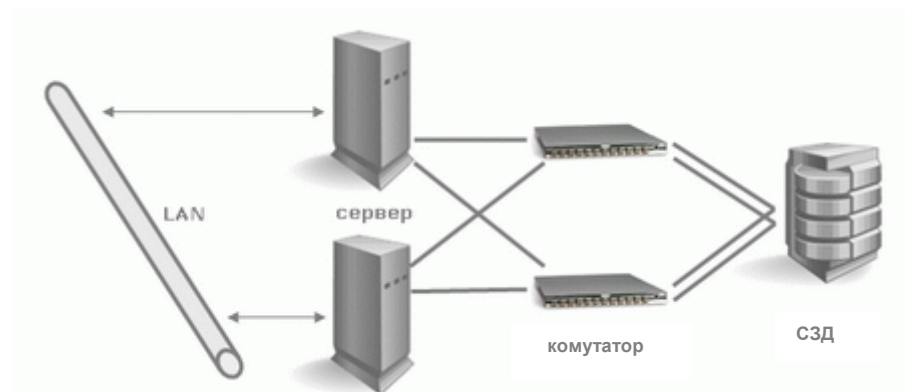


Рисунок 4.5

На рис. 4.5 приведено приклад високонадійної інфраструктури, в якій сервери включені одночасно у локальну мережу (ліворуч) і в мережу зберігання даних (праворуч). Така схема забезпечує доступ до даних, що знаходяться на СЗД, при виході з ладу будь-якого процесорного модулю, комутатору або шляху доступу.

Використання SAN дозволяє забезпечити:

- централізоване управління ресурсами серверів і систем зберігання даних;

- підключення нових дискових масивів і серверів без зупинки роботи всієї системи зберігання;
- використання раніше придбаного обладнання разом з новими пристроями зберігання даних;
- оперативний і надійний доступ до накопичувачів даних, що знаходяться на великій відстані від серверів, без значних втрат продуктивності;
- прискорення процесу резервного копіювання та відновлення даних - BURA.

Архітектура SAN

SAN є високошвидкісною мережею передавання даних, яка призначена для підключення серверів до пристроїв зберігання даних. Різні топології SAN (точка-точка, петля з арбітражною логікою (Arbitrated Loop) і комутація) замінюють традиційні шинні з'єднання «сервер - пристрої зберігання» і надають у порівнянні з ними велику гнучкість, продуктивність та надійність. В основі концепції SAN лежить можливість з'єднання будь-якого серверу з будь-яким пристроєм зберігання даних, що працює по протоколу Fibre Channel. Принцип взаємодії вузлів у SAN з топологіями точка-точка (рис. 4.8) або комутацією показаний на рис. 4.6, рис. 4.7. В SAN з топологією Arbitrated Loop передавання даних здійснюється послідовно від вузла до вузла. Для того, щоб почати передавання даних передавальний пристрій ініціалізує арбітраж за право використання середовища передавання даних (звідси і назва топології - Arbitrated Loop).

Транспортну основу SAN складає протокол Fibre Channel, який використовує як мідні, так і волоконно-оптичні з'єднання пристроїв.

Компоненти SAN

Компоненти SAN поділяються на наступні:

- Host Bus Adaptors (HBA);
- Ресурси зберігання даних;
- Пристрої, що реалізують інфраструктуру SAN;
- Програмне забезпечення.

Петля з арбітражною логікою

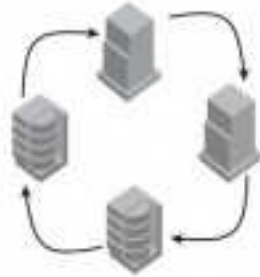


Рисунок 4.6 - Host Bus Adaptors

НБА встановлюються у сервери і здійснюють їх взаємодію з SAN по протоколу Fibre Channel. Стек протоколів Fibre Channel реалізований всередині НБА. Найбільш відомими виробниками НБА є компанії Emulex, JNI, Qlogic і Agilent.

Комутація

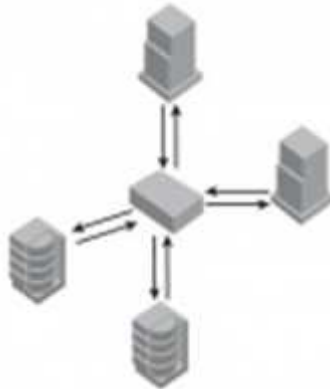


Рисунок 4.7 - Ресурси зберігання даних

До ресурсів зберігання даних відносяться дискові масиви, стрічкові накопичувачі та бібліотеки з інтерфейсом Fibre Channel. Багато своїх можливостей ресурси зберігання реалізують тільки тоді, коли вони включені в SAN. Так дискові масиви вищого класу можуть здійснювати реплікацію даних між масивами по мережах Fibre Channel, а стрічкові бібліотеки можуть реалізовувати перенесення даних на стрічку прямо з дискових масивів з

інтерфейсом Fibre Channel, минаючи мережу і сервери (Serverless backup). Найбільшу популярність на ринку придбали дискові масиви компаній EMC, Hitachi, IBM, Compaq (сімейство Storage Works), а з виробників стрічкових бібліотек слід згадати StorageTek, Quantum/ATL, IBM.

Точка-точка

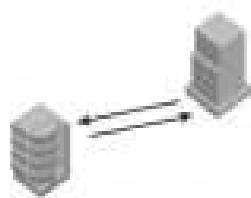


Рисунок 4.8

Пристрої, що реалізують інфраструктуру SAN

Пристроями, що реалізують інфраструктуру SAN, є комутатори Fibre Channel (Fibre Channel switches, FC switches), концентратори (Fibre Channel Hub) і маршрутизатори (Fibre Channel-SCSI routers). Концентратори використовуються для об'єднання пристроїв, що працюють в режимі Fibre Channel Arbitrated Loop (FC_AL). Застосування концентраторів дозволяє підключати і відключати пристрої в петлі без зупинки системи, оскільки концентратор автоматично замикає петлю у випадку відключення пристрою і автоматично розмикає петлю, якщо до нього був підключений новий пристрій. Кожна зміна петлі супроводжується складним процесом її ініціалізації. Процес ініціалізації багатоступеневий, і до його закінчення обмін даними в петлі неможливий.

Усі сучасні SAN побудовані на комутаторах, що дозволяють реалізувати повноцінне мережне з'єднання. Комутатори можуть не тільки з'єднувати пристрої Fibre Channel, а й розмежовувати доступ між пристроями, для чого на комутаторах створюються так названі зони. Пристрої, поміщені у різні зони, не можуть обмінюватися інформацією один з одним. Кількість портів у SAN можна збільшувати, з'єднуючи комутатори один з одним. Група зв'язаних

комутаторів називається Fibre Channel Fabric або просто Fabric. Зв'язки між комутаторами називають Interswitch Links або скорочено ISL.

Програмне забезпечення

Програмне забезпечення дозволяє реалізувати резервування шляхів доступу серверів до дискових масивів і динамічний розподіл навантаження між шляхами. Для більшості дискових масивів існує простий спосіб визначити, що порти, які доступні через різні контролери, відносяться до одного диску. Спеціалізоване програмне забезпечення підтримує таблицю шляхів доступу до пристроїв і забезпечує відключення шляхів у випадку аварії, динамічне підключення нових шляхів і розподіл навантаження між ними. Як правило, виробники дискових масивів пропонують спеціалізоване програмне забезпечення такого типу для своїх масивів. Компанія VERITAS Software виробляє програмне забезпечення VERITAS Volume Manager, яке призначене для організації логічних дискових томів з фізичних дисків і забезпечує резервування шляхів доступу до дисків, а також розподіл навантаження між ними для більшості відомих дискових масивів.

Протоколи, які використовуються

У мережах зберігання даних використовуються низькорівневі протоколи, такі як Fibre Channel Protocol (FCP), транспорт SCSI через Fibre Channel – протокол, який використовується найчастіше. Існує у варіантах 1 Gbit/s, 2 Gbit/s, 4 Gbit/s, 8 Gbit/s і 10 Gbit/s; iSCSI, транспорт SCSI через TCP/IP; FCoE, транспортування FCP/SCSI поверх "чистого" Ethernet; FCIP і iFCP, інкапсуляція і передавання FCP/SCSI у пакетах IP; HyperSCSI, транспорт SCSI через Ethernet; FICON транспорт через Fibre Channel (використовується тільки мейнфреймами); ATA over Ethernet, транспорт ATA через Ethernet; SCSI і/або TCP/IP транспорт через InfiniBand (IB).

Переваги

Висока надійність доступу до даних, що знаходяться на зовнішніх системах зберігання. Незалежність топології SAN від СЗД і серверів, які використовуються.

Централізоване зберігання даних (надійність, безпека); зручне централізоване управління комутацією і даними; перенесення інтенсивного трафіку введення-виведення в окрему мережу - розвантаження LAN; висока швидкодія та низька латентність; масштабованість і гнучкість логічної структури SAN; географічні розміри SAN, на відміну від класичних DAS, практично не обмежені; можливість оперативно розподіляти ресурси між серверами; можливість будувати відмовостійкі кластерні рішення без додаткових витрат на базі SAN, яка є в наявності; проста схема резервного копіювання - всі дані знаходяться в одному місці; наявність додаткових можливостей і сервісів (снапшоти, віддалена реплікація); висока степінь безпеки SAN.

Спільне використання систем зберігання як спрощує адміністрування і додає значну гнучкість, оскільки кабелі та дискові масиви не потрібно фізично транспортувати і перекомутувати від одного сервера до іншого.

Іншою перевагою є можливість завантажувати сервери прямо з мережі зберігання. При такій конфігурації можна швидко і легко замінити сервер, який дає збій, переконфігурувавши SAN таким чином, що сервер-заміна буде завантажуватися з LUN'a сервера, який дав збій. Ця процедура може зайняти, наприклад, півгодини. Ідея відносно нова, але вже використовується у новітніх датацентрах.

Також мережі зберігання допомагають більш ефективно відновлювати працездатність після збою. У SAN може входити віддалена ділянка із вторинним пристроєм зберігання. У такому випадку можна використовувати реплікацію - реалізовану на рівні контролерів масивів, або за допомогою спеціальних апаратних пристроїв. Оскільки канали WAN на основі протоколу IP зустрічаються часто, були розроблені протоколи Fibre Channel over IP (FCIP) і iSCSI з метою розширити єдину SAN засобами мереж на основі протоколу IP.

Недоліки

Всі недоліки зводяться тільки до високої вартості подібних рішень. Український ринок СЗД в цілому відстає від ринку західних розвинених країн,

особливо - в широкому використанні мереж зберігання даних. Зокрема, певний вплив продовжують надавати дефіцит і висока вартість швидкісних каналів зв'язку.

Відмінність від NAS

Основна відмінність між SAN і NAS полягає у способі організації обміну даними між пристроями зберігання і серверами. Взагалі кажучи, архітектура SAN направлена на вирішення проблем, що викликаються інтенсивними процедурами резервного копіювання та обміну даними шляхом перенесення всієї системи у виділену підмережу. Основані на протоколі Fibre Channel системи SAN дозволяють у широких межах змінювати ємність системи зберігання даних і гарантувати більш високу пропускну спроможність у межах виділеної підмережі (дискові масиви та стрічкові бібліотеки, не обладнані інтерфейсами Fibre Channel, можна підключити до SAN, використовуючи маршрутизатори Fibre Channel-SCSI).

Варіанти організації доступу до сторедж-систем

Розрізняють три основних варіанти організації доступу до систем зберігання:

- SAS (Server Attached Storage) – сторедж, приєднаний до сервера;
- NAS (Network Attached Storage) – сторедж, приєднаний до мережі;
- SAN (Storage Area Network) – мережа зберігання даних.

Розглянемо топології відповідних сторедж-систем та їх особливості.

SAS

Сторедж-система, приєднана до сервера (рис. 4.9). Знайомий усім, традиційний спосіб підключення системи зберігання даних до високошвидкісного інтерфейсу у сервері, як правило, до паралельного SCSI інтерфейсу.

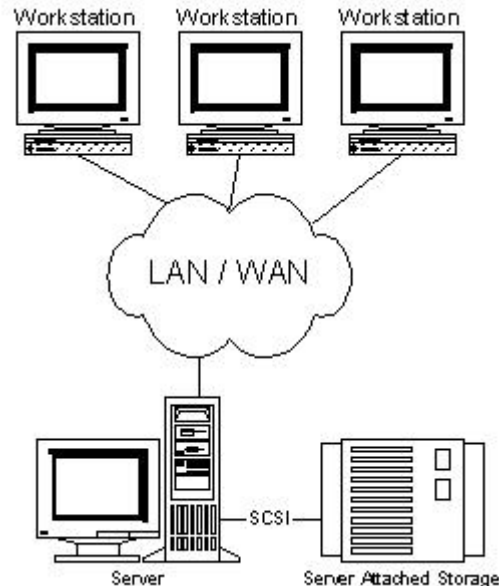


Рисунок 4.9 - Server Attached Storage

Використання окремого корпусу для сторедрж-системи в межах топології SAS не є обов'язковим. Основна перевага сторедрж, приєднаного до сервера, порівняно з іншими варіантами - низька ціна і висока швидкодія з розрахунку один сторедрж для одного сервера. Така топологія є найоптимальнішою в разі використання одного сервера, через який організовується доступ до масиву даних. Але в неї залишається ряд проблем, які спонукали проектувальників шукати інші варіанти організації доступу до систем зберігання даних.

До особливостей SAS можна віднести: доступ до даних залежить від ОС і файлової системи (у загальному випадку); складність організації систем з високою готовністю; низька вартість; висока швидкодія в межах однієї ноди; зменшення швидкості відгуку при завантаженні сервера, який обслуговує сторедрж.

NAS

Сторедж-система, під'єднана до мережі (рис. 4.10). Цей варіант організації доступу з'явився порівняно недавно. Основною його перевагою є зручність інтеграції додаткової системи зберігання даних в існуючі мережі, але сам по собі він не привносить хоч трохи радикальних покращень в архітектуру сторедрж. Фактично NAS є чистим файл-сервером, і сьогодні можна зустріти

чимало нових реалізацій сторедж типу NAS на основі технології тонкого сервера (Thin Server).

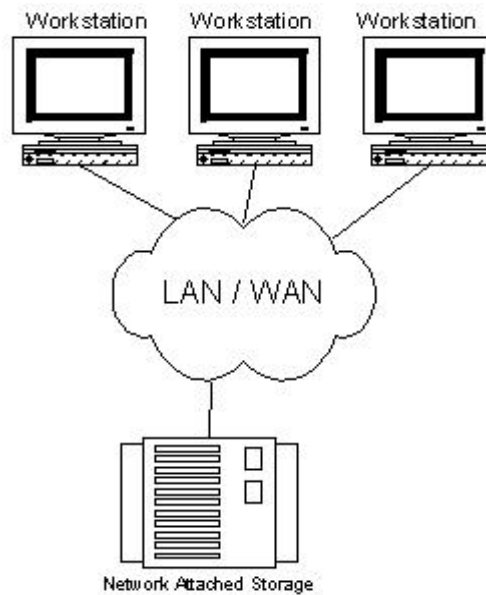


Рисунок 4.10 - Network Attached Storage

Особливості NAS: виділений файл-сервер; доступ до даних не залежить від ОС і платформи; зручність адміністрування; максимальна простота установки; низька масштабованість; конфлікт з трафіком LAN/WAN. Сторедж, побудований по технології NAS, є ідеальним варіантом для дешевих серверів з мінімальним набором функцій.

SAN

Основою SAN є окрема від LAN/WAN мережа, яка використовується для організації доступу до даних серверів і робочих станцій, що займаються їх прямою обробкою (рис. 4.11). Така мережа створюється на основі стандарту Fibre Channel, що дає сторедж-системам переваги технологій LAN/WAN і можливості по організації стандартних платформ для систем з високою готовністю і високою інтенсивністю запитів. Майже єдиним недоліком SAN на сьогодні залишається відносно висока ціна компонент, але при цьому загальна вартість володіння для корпоративних систем, побудованих з використанням технології мереж зберігання даних, є досить низькою.

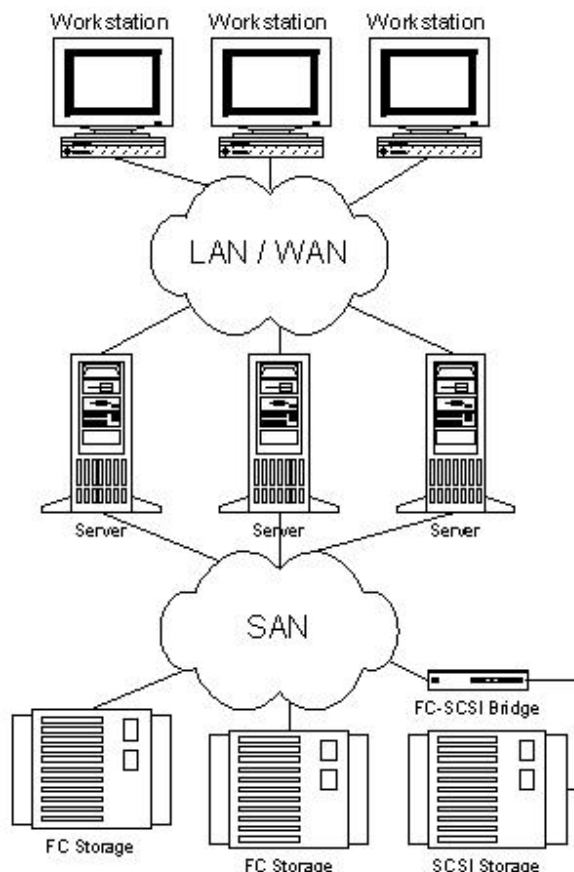


Рисунок 4.11 - Storage Area Network

До основних переваг SAN можна віднести практично всі її особливості: незалежність топології SAN від стореждж-систем і серверів; зручне централізоване управління; відсутність конфлікту з трафіком LAN/WAN; зручне резервування даних без завантаження локальної мережі і серверів; висока швидкодія; висока масштабованість; висока гнучкість; висока готовність і відмовостійкість.

Необхідно також зазначити, що технологія ця ще досить молода і найближчим часом вона повинна пережити чимало удосконалень в області стандартизації управління і способів взаємодії SAN підмереж.

FC як основа побудови SAN

Подібно LAN, SAN може створюватися з використанням різних топологій і носіїв. При побудові SAN може використовуватися як паралельний SCSI інтерфейс, так і Fibre Channel або SCI (Scalable Coherent Interface), але своєю популярністю SAN зобов'язана саме Fibre Channel. У проектуванні цього інтерфейсу брали участь фахівці зі значним досвідом у розробці як каналних,

так і мережних інтерфейсів, і їм вдалося об'єднати всі важливі позитивні риси обох технологій для того, щоб отримати щось справді революційно нове.

Основні ключові особливості каналних інтерфейсів: низькі затримки; високі швидкості; висока надійність; топологія точка-точка; невеликі відстані між нодами; залежність від платформи. Основні ключові особливості мережних інтерфейсів: багатоточкові топології; великі відстані; висока масштабованість; низькі швидкості; програмна завантаження; великі затримки. Дані особливості об'єдналися у Fibre Channel: високі швидкості; незалежність від протоколу (0-3 рівні); великі відстані; низькі затримки; висока надійність; висока масштабованість; багатоточкові топології.

Традиційно сторедж інтерфейси (те, що знаходиться між хостом і пристроями зберігання інформації) були перешкодою на шляху до зростання швидкодії і збільшення обсягу систем зберігання даних. У той же час прикладні задачі вимагають значного приросту апаратних потужностей, які, в свою чергу, тягнуть за собою потребу у збільшенні пропускної спроможності інтерфейсів для зв'язку з сторедж-системами. Саме проблеми побудови гнучкого високошвидкісного доступу до даних допомагає вирішити Fibre Channel.

На протязі двох років була проведена велика робота по узгодженню взаємодії виробників різних компонент, і було зроблено все необхідне, щоб Fibre Channel перетворився з чисто концептуальної технології в реальну, яка отримала підтримку у вигляді інсталяцій в лабораторіях і обчислювальних центрах. Були спроектовані перші комерційні зразки компонентів для побудови SAN на базі FC, таких як адаптери, хаби, свічі і мости. На сьогоднішній день FC використовується у комерційних цілях у діловій сфері, на виробництві і в масштабних проектах реалізації систем, які критичні до відмов.

Fibre Channel - це відкритий промисловий стандарт високошвидкісного послідовного інтерфейсу. Він забезпечує підключення серверів і сторедж-систем на відстані до 10 км (при використанні стандартного оснащення) на швидкості 100 – 200 MB/s. Fibre Channel одночасно підтримує ряд стандартних протоколів (серед яких TCP/IP і SCSI-3) при використанні одного фізичного

носія, який потенційно спрощує побудову мережної інфраструктури, до того ж це надає можливості для зменшення вартості монтажу та обслуговування. Проте використання окремих підмереж для LAN/WAN і SAN має ряд переваг і є рекомендованим за замовчуванням.

Однією з найважливіших переваг Fibre Channel поряд із швидкісними параметрами (які, до речі, не завжди є основними для користувачів SAN і можуть бути реалізовані за допомогою інших технологій) є можливість роботи на великих відстанях і гнучкість топології, яка прийшла у новий стандарт із мережних технологій. Таким чином, концепція побудови топології мережі зберігання даних базується на тих же принципах, що і традиційні мережі, як правило, на основі концентраторів і комутаторів (станом на сьогоднішній день – комутаторів), які допомагають запобігти падінню швидкості при зростанні кількості нод і створюють можливості зручної організації систем без єдиної точки відмов.

Для кращого розуміння переваг і особливостей цього інтерфейсу наведемо порівняльну характеристику FC і Parallel SCSI у табл. 4.6.

Таблиця 4.6 – Порівняння технологій Fiber Channel та паралельного SCSI

	Fibre Channel	Parallel SCSI
Швидкодія	100MB/sec Новий стандарт: 200MB/sec & 400MB/sec	Ultra 160m – 160MB/sec Новий стандарт: 320MB/sec
Максимальна відстань	Copper: 30m Fiber optic: 2-10km	Copper, single-ended: 3m Copper, differential: 25m
Протоколи, які підтримуються	SCSI, TCP/IP, VI, IPI, ESCON, HIPPI, FCON	SCSI
Максимальна кількість пристроїв	127 на кільце, 2 ²⁴ на комутатор	16 на канал
Топології	кільце, точка-точка, комутатор	точка-точка, складна реалізація пристрою комутації каналів

У стандарті Fibre Channel передбачається використання різноманітних топологій, таких як точка-точка (Point-to-Point), кільце або FC-AL концентратор (Loop або Hub FC-AL), магістральний комутатор (Fabric/Switch). Топологія point-to-point використовується для під'єднання одиночної сторедж-системи до сервера.

Loop або Hub FC-AL - для під'єднання множинних сторедж пристроїв до декількох хостів. При організації подвійного кільця збільшується швидкодія і відмовостійкість системи. Комутатори використовуються для забезпечення максимальної швидкодії і відмовостійкості для складних, великих і розгалужених систем. Завдяки мережній гнучкості у SAN закладена надзвичайно важлива особливість - зручна можливість побудови відмовостійких систем. Пропонуючи альтернативні рішення для систем зберігання даних і можливості по об'єднанню декількох сторедж для резервування апаратних засобів, SAN допомагає забезпечувати захист апаратно-програмних комплексів від апаратних збоїв. Для демонстрації наведемо приклад створення двухнодової системи без точок відмов (рис. 4.12).

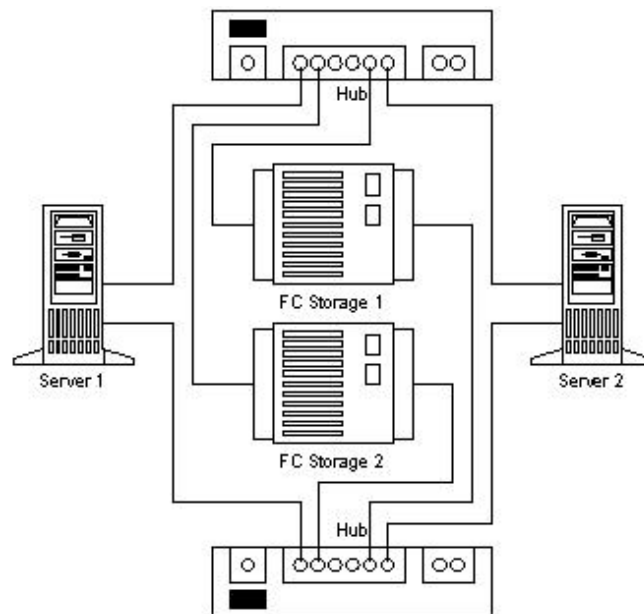


Рисунок 4.12 - No Single Point of Failure

Побудова трьох- і більше нодових систем здійснюється простим додаванням в FC мережу додаткових серверів і підключенням їх до обох

концентраторів/ комутаторів). При використанні FC побудова стійких до збоїв (disaster tolerant) систем стає прозорою. Мережні канали і для стореж, і для локальної мережі можна прокласти на основі оптоволокна (до 10 км і більше з використанням підсилювачів сигналу) як фізичного носія для FC, при цьому використовується стандартна апаратура, яка дає можливість значно зменшити вартість таких систем.

Завдяки можливості доступу до всіх компонентів SAN з будь-якої її точки ми отримуємо надзвичайно гнучку управляєму мережу даних. При цьому слід зауважити, що в SAN забезпечується прозорість (можливість бачити) всі компоненти аж до дисків у стореж-системах. Ця особливість підштовхнула виробників компонентів до використання свого значного досвіду у побудові систем управління для LAN/WAN з тим, щоб закласти широкі можливості з моніторингу та управління у всі компоненти SAN. Ці можливості включають в себе моніторинг і управління окремих нод, стореж компонентів, корпусів, мережних пристроїв і мережних підструктур.

У системі управління і моніторингу SAN використовуються такі відкриті стандарти, як:

- SCSI command set;
- SCSI Enclosure Services (SES);
- SCSI Self Monitoring Analysis and Reporting Technology (S.M.A.R.T.);
- SAF-TE (SCSI Accessed Fault-Tolerant Enclosures);
- Simple Network Management Protocol (SNMP);
- Web-Based Enterprise Management (WBEM).

Системи, які побудовані з використанням технологій SAN, не тільки забезпечують адміністратору можливість стежити за розвитком і станом стореж ресурсів, а й відкривають можливості з моніторингу та контролю трафіку. Завдяки таким ресурсам програмні засоби управління SAN реалізують найбільш ефективні схеми планування обсягу стореж і балансування навантаження на компоненти системи. Мережі зберігання даних добре інтегруються в існуючі інформаційні інфраструктури. Їх впровадження не

вимагає яких-небудь змін у вже існуючих мережах LAN і WAN, а лише розширює можливості існуючих систем, позбавляючи їх від задач, орієнтованих на передавання великих обсягів даних. Причому при інтеграції і адмініструванні SAN дуже важливим є те, що ключові елементи мережі підтримують гарячу заміну і установку, з можливостями динамічного конфігурування. Так що додати той чи інший компонент або здійснити його заміну адміністратор може, не вимикаючи систему. І весь цей процес інтеграції може бути візуально відображений у графічній системі управління SAN.

Розглянувши перераховані вище переваги, можна виділити ряд ключових моментів, які безпосередньо впливають на одну з основних переваг Storage Area Network - загальну вартість володіння (Total Cost Ownership). Неймовірні можливості масштабування дозволяють підприємству, яке використовує SAN, вкладати кошти у сервери і сторедж по мірі необхідності. А також зберегти свої вкладення у вже інсталювану техніку при зміні технологічних поколінь. Кожний новий сервер буде мати можливість високошвидкісного доступу до сторедж і кожний додатковий гігабайт сторедж буде доступний усім серверам підмережі по команді адміністратора.

Можливості по побудові відмовостійких систем можуть приносити пряму комерційну вигоду від мінімізації простоїв і рятувати систему у випадку виникнення стихійного лиха або інших катаклізмів. Управляємість компонентів і прозорість системи надають можливість здійснювати централізоване адміністрування усіх сторедж ресурсів, а це, в свою чергу, значно зменшує витрати на їх підтримку, вартість якої, як правило, становить понад 50% від вартості обладнання.

Вплив SAN на прикладні задачі

Наведемо кілька прикладів прикладних задач, які без використання мереж зберігання даних вирішувалися б неефективно, вимагали б великих фінансових вкладень або ж взагалі не вирішувалися б стандартними методами.

Резервування і відновлення даних (Data Backup and Recovery).

Використовуючи традиційний SCSI інтерфейс, перед користувачем при побудові систем резервування і відновлення даних виникає ряд складних проблем, які можна дуже просто вирішити, використовуючи технології SAN і FC (табл. 4.7). Використання мереж зберігання даних виводить вирішення задачі резервування і відновлення на новий рівень і надає можливість здійснювати бекап у кілька разів швидше, ніж раніше, без завантаження локальної мережі і серверів роботою по резервуванню даних.

Таблиця 4.7

Проблеми (Data Backup and Recovery)	Рішення з використанням технологій SAN і FC
Завантаження LAN	LAN-free Backup (окрема сторедж мережа)
Завантаження серверів	Serverless Backup (storage to storage)
Проблема віддаленого резервування	Віддалення сторедж на відстань > 10km через FC з використанням FC/WAN Gateway
Мале вікно (відсутність часу) для backup	Висока пропускну спроможність мережі FC

Кластеризація серверів (Server Clustering).

Однією з типових задач, для яких ефективно використовується SAN, є кластеризація серверів. Оскільки одним із ключових моментів є організація високошвидкісних кластерних систем, які працюють з даними - це доступ до сторедж, то з появою SAN побудова багатонодових кластерів на апаратному рівні вирішується простим додаванням сервера з підключенням до SAN (це можна зробити, навіть не вимикаючи системи, оскільки свічі FC підтримують hot-plug). При використанні паралельного SCSI інтерфейсу, можливості по приєднанню і масштабованість якого значно гірше, ніж у FC, кластери, орієнтовані на обробку даних, було б важко зробити з кількістю нод більше двох. Комутатори паралельного SCSI - досить складні і дорогі пристрої, а для

FC це стандартний компонент. Для створення кластеру, який не матиме жодної точки відмов, досить інтегрувати у систему дзеркальну SAN (технологія DUAL Path).

У межах кластеризації одна з технологій RAIS (Redundant Array of Inexpensive Servers) є актуальною для побудови потужних масштабованих систем інтернет-комерції та інших видів завдань з підвищеними вимогами до потужності. За словами Alistair A. Croll, співзасновника Networkshop Inc, використання RAIS виявляється досить ефективним: «Наприклад, за \$ 12000-15000 ви можете купити близько шести недорогих одно-двопроцесорних (Pentium III) Linux/Apache серверів. Потужність, масштабованість і відмовостійкість такої системи буде значно вищою, ніж, наприклад, у одного чотирьохпроцесорного серверу на базі процесорів Xeon, а вартість однаковою».

Одночасний доступ до відео і розподіл даних (Concurrent video streaming, data sharing).

Представимо собі задачу, коли нам потрібно на декількох (скажімо, >5) станціях редагувати відео або просто працювати над даними великого обсягу. Передавання файлу розміром 100GB по локальній мережі забере у нас кілька хвилин, а загальна робота над ним буде дуже складним завданням. При використанні SAN кожна робоча станція і сервер мережі отримують доступ до файлу на швидкості, еквівалентній локальному високошвидкісного диску. Якщо нам потрібні ще одна станція/сервер для обробки даних, ми зможемо її додати до SAN, не вимикаючи мережі, простим під'єднанням станції до SAN комутатора і наданням їй прав доступу до сторедж. Якщо ж нас перестане задовольняти швидкодія підсистеми даних, ми зможемо просто додати ще один сторедж і з використанням технології розподілу даних (наприклад, RAID 0) отримати вдвічі більшу швидкодію.

Основні компоненти SAN

Середовище. Для з'єднання компонентів у рамках стандарту Fibre Channel використовують мідні та оптичні кабелі. Обидва типи кабелів можуть використовуватися одночасно при побудові SAN. Конверсія інтерфейсів

здійснюється за допомогою GBIC (Gigabit Interface Converter) і MIA (Media Interface Adapter). Обидва типи кабелю на сьогодні забезпечують однакову швидкість передавання даних. Мідний кабель використовується для коротких відстаней (до 30 метрів), оптичний - як для коротких, так і для відстаней до 10 км і більше. Використовують багатомодовий і одномодовий оптичні кабелі. Багатомодовий (Multimode) кабель використовується для коротких відстаней (до 2 км). Внутрішній діаметр оптоволокна мультимодового кабелю становить 62,5 або 50 мікронів. Для забезпечення швидкості передавання 100 МБ/с (200 МБ/с в дуплексі) при використанні багатомодового оптоволокна довжина кабелю не повинна перевищувати 200 метрів. Одномодовий кабель використовується на великі відстані. Довжина такого кабелю обмежена потужністю лазера, який використовується у передавачі сигналу. Внутрішній діаметр оптоволокна одномодового кабелю становить 7 або 9 мікрон, він забезпечує проходження одиночного променя.

Конектори, адаптери. Для під'єднання мідних кабелів використовуються конектори типу DB-9 або HSSD. HSSD вважається більш надійним, але DB-9 використовується так само часто, тому що він більш простий і дешевий. Стандартним (найбільш поширеним) конектором для оптичних кабелів є SC конектор, він забезпечує якісне, чітке з'єднання. Для звичайного підключення використовуються багатомодові SC конектори, а для віддаленого - одномодові. У багатопортових адаптерах використовуються мікроконектори.

Найбільш розповсюджені адаптери для FC під шину PCI 64 bit. Також багато FC адаптерів виробляється під шину S-BUS, для спеціалізованого використання випускаються адаптери під MCA, EISA, GIO, HIO, PMC, Compact PCI. Найрозповсюдженіші - однопортові, зустрічаються двох- і чотирьохпортова картки. На PCI адаптери, як правило, використовують DB-9, HSSD, SC конектори. Також часто зустрічаються GBIC-based адаптери, які поставляються як з модулями GBIC, так і без них. Fibre Channel адаптери відрізняються класами, які вони підтримують, і різноманітними особливостями.

Для розуміння відмінностей наведемо порівняльну таблицю адаптерів виробництва фірми QLogic (табл. 4.8).

Таблиця 4.8

Fibre Channel Host Bus Adapter Family Chart														
SANblade	64 Bit	FCAL Publ. Pvt Loop	FL Port	Class 3	F Port	Class 2	Point to Point	IP/ SCSI	Full Duplex	FC Tape	PCI 1.0 Hot Plug Spec	Solaris Dynamic Reconfig	VIB	2Gb
2100 Series	33 & 66MHz PCI	X	X	X										
2200 Series	33 & 66MHz PCI	X	X	X	X	X	X	X	X	X				
	33MHz PCI	X	X	X	X	X	X	X	X	X	X			
	25 MHz Sbus	X	X	X	X	X	X	X	X	X		X		
2300 Series	66 MHz PCI/ 133MHz PCI- X	X	X	X	X	X	X	X	X	X			X	X

Комутатори. Fibre Channel Switches (комутатори) мають ті самі функції, що й звичайні LAN комутатори. Вони забезпечують повношвидкісне неблоковане підключення між нодами. Будь-яка нода, що підключена до FC комутатору, отримує повну (з можливостями масштабування) смугу пропускання. При збільшенні кількості портів комутованої мережі її пропускна спроможність збільшується. Комутатори можуть використовуватися разом із концентраторами (які використовують для ділянок, що не вимагають виділеної смуги пропускання для кожної ноди) для досягнення оптимального співвідношення ціна/продуктивність. Завдяки каскадуванню свічі потенційно можуть використовуватися для створення FC мереж з кількістю адрес 2^{24} (понад 16 мільйонів).

Сервери та Сторедж. Особливу увагу потрібно приділяти особливостям реалізації компонент виробниками SAN і програмних засобів управління.

На сьогоднішній день Storage Area Network є досить новою технологією, яка незабаром може стати масовою у колі корпоративних

замовників. У Європі та США підприємства, які мають досить великий парк інсталюваних стореждж-систем, вже починають переходити на мережі зберігання даних для організації стореждж з найкращим показником загальної вартості володіння.

За прогнозами аналітиків, значна кількість серверів середнього та верхнього рівня будуть поставлятися з попередньо встановленим інтерфейсом Fibre Channel (таку тенденцію можна помітити вже сьогодні), і лише для внутрішнього підключення дисків у серверах буде використовуватися паралельний SCSI інтерфейс. Вже сьогодні при побудові стореждж-систем і придбанні серверів середнього та верхнього рівня слід звернути увагу на дану перспективну технологію, тим більше, що вже сьогодні вона дає можливість реалізувати ряд завдань значно дешевше, ніж за допомогою спеціалізованих рішень. Крім того, вкладаючи у технологію SAN сьогодні, ми не втрачаємо свої вкладення завтра, оскільки особливості Fibre Channel створюють прекрасні можливості для використання у майбутньому вкладених сьогодні інвестицій.

Література

1. Иванов И.В. Теория информационных процессов и систем: Учебное пособие. — 2-е изд., перераб. и доп. — Белгород: БГТУ, 2014. — 203 с.
2. Галашев В.А. Системы поиска и обработки информации: Учеб.-метод. пособие. Ижевск: Удм. гос. ун-т., 2011. — 149 с.
3. Ландэ Д.В. Поиск знаний в Internet. Профессиональная работа.: Пер.с англ. — М.: Издательский дом «Вильямс», 2005.-272 с.
4. Алексеев Е.Г., Богатырев С.Д. Информатика. Мультимедийный электронный учебник.
5. Аликберов А., Базак Д., Колмановская Е., Ланде Д. Поисковые системы Интернет. Мультимедийный электронный учебник. [<http://bourabai.kz/dbt/seo/searchsystems.htm>]
6. Антон Кириллов «Поисковые системы изнутри». Мультимедийный электронный учебник. [<https://compress.ru/article.aspx?id=21155>]
7. Михаил Сочин «100 самых важных факторов ранжирования и критериев для продвижения веб-сайта в поисковых системах Google и Яндекс». Интернет ресурс. [<https://www.seo-website.ru/100-seo-factors-of-ranking-in-google/>]