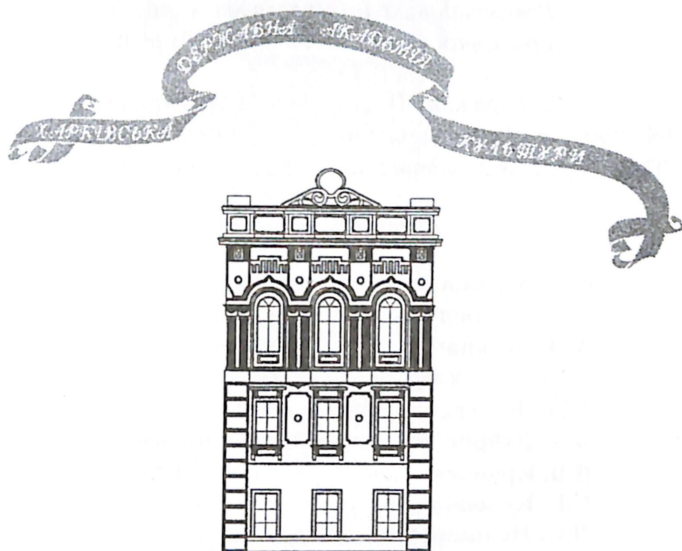


МІНІСТЕРСТВО КУЛЬТУРИ І МИСТЕЦТВ УКРАЇНИ

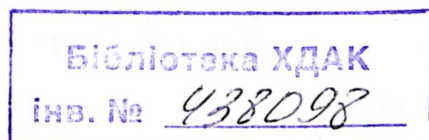


ВІСНИК

Харківської державної академії культури

Випуск 3

Збірник наукових праць



Харків- ХДАК-2000

установ. Основою інформаційної взаємодії у всі часи і у всіх народів було і залишається міжособове спілкування, співробітництво людей.

Підсумовуючи, зазначимо, що свобода функціонування інформації, забезпечення її доступності, зберігання конфіденційності, дотримання міри у використанні та ін. є основними морально-етичними принципами демократичного суспільства, базою формування інформаційної цивілізації. Використання їх у повсякденному житті - показник рівня інформаційної культури, досягнутої світовим суспільством, державою або конкретною особою.

Зауважимо, що за межами статті залишилась проблема міжособової інформаційної взаємодії. Вона є самостійним предметом вивчення і потребує додаткових уточнень. Зважаючи на її багатогранність, хотілося б запросити всіх зацікавлених взяти участь у її вирішенні.

СПИСОК ЛІТЕРАТУРИ

1. Закон України "Про інформацію" // Відом. Верх. Ради України. - 1992. - № 48. - Ст. 1447-1462.
2. Закон України "Про науково-технічну інформацію" // Відом. Верх. Ради України. - 1993. - № 33. - Ст. 843-851.
3. Библер В.С. Культура. Диалог культур // Вопр. филос. - 1989. - № 6. - С. 31-42.
4. Уледова И. Моральная ответственность и нравственная культура // Библиотекарь. - 1982. - № 9. - С. 30-31.
5. Федоренко Е.М. Профессиональная этика. - К.: Вища школа, 1991. - 190 с.
6. Watson David. Computer confidentiality and privation // Comput. Hum. Serv. - 1989. - № 1-2. - P. 153-168.

Надійшла до редколегії 3.04.2000 р.

УДК 681.3.06

Б.М. СМОЛЯНИЦЬКИЙ

КОДУВАННЯ ТЕКСТОВОЇ ІНФОРМАЦІЇ В ІНТЕРНЕТІ

З системних позицій розглянуто проблему кодування різноманітних символів в Інтернеті і можливі шляхи її вирішення з метою підвищення ефективності використання глобальної мережі в документно-комунікаційній сфері

Проаналізуємо методологічні та технологічні особливості методів кодування інформації в Інтернеті.

Текстові файли – найпоширеніший тип даних не лише в Інтернеті, а й в

усьому комп'ютерному світі. Існують два дуже важливі чинники їх системного аналізу. Перший – надзвичайно велика кількість символів, потрібних для підтримки різноманітних мов. Американські програмісти для роботи зі 128 символами використовують набір символів US ASCII. ASCII – аббревіатура від англ. American Standard Code for Information Interchange Американський стандартний код для інформаційних обмінів. ASCII – це угода про стандарт подання літер, цифр, символів та ін. Важливо пам'ятати, що більше 250 символів необхідні лише для того, щоб охопити пару десятків європейських мов, що базуються на латинському алфавіті. Для підтримки інших алфавітів (кирилиці, грецького, івриту, арабського та ін.) додатково потрібно більше сотні символів; китайська, японська і корейська мови додають до цього списку понад десять тисяч ієрогліфів. Часи, коли в Інтернеті були переважно англومовні користувачі, минають. Гнучкі розробники програмного забезпечення використали це для створення багатомовних програм.

Другий чинник характеризується тим, що суто текстові дані трапляються істотно рідше. Користувачі Інтернету прагнуть, щоб документи, які друкуються, містили графіки, діаграми, примітки, заголовки і при цьому використовувалися різноманітні шрифти. Документи з мережі можуть вмістити мультимедіацію, посилання на мережеві бази даних і звуковий супровід. Внаслідок комбінацій різноманітних типів даних утворюються документи мультимедіа. Текстовий формат (звичайно використовується як базовий) є стартовою точкою багатьох форматів документів мультимедіа. Багато з тих форматів в Інтернеті – не просто текстові формати, коректніше їх слід було б називати форматами документа. Такі формати забезпечують комбінування текстових, графічних та інших форм даних у процесі розробки документів.

Детальніше розглянемо характеристику наборів символів, які є основою означених документів.

Якщо критично проаналізувати різні матеріали про символи і набори символів, напевно можна дійти висновку, що спиратися на поняття символ дуже складно. Воно має різні відтінки значень, тому слід уникати слів символ і набір символів у разі розбіжності. Спеціалісти в галузі інформаційних технологій вважають, що термін набір символів є помилковим, оскільки припускає розбіжність значень. [2].

Більшість користувачів упевнені в тому, що **A**, **A** й **А** – це один символ, незважаючи на те, що на вигляд вони різні. Для уточнення зовнішнього вигляду символу типографи використовують термін гліф (glyph). Так, незважаючи на те, що всі гліфи **A**, **A**, **А** це той самий символ, важко помітити, що вони різні. Точніше, гліф – це специфічне візуальне подання символу.

Звичайно, мало кого цікавить вигляд окремо взятого символу чи гліфа. Потрібно, щоб текст мав гарний вигляд, тобто є сенс говорити про підбір символів. Навіть в американському англійському він повинен мати п'ятдесят два великі та малі символи, десять цифр і різні знаки пунктуації. Такий підбір називають репертуаром (repertoire). Відповідний набір гліфів, по одному для кожного символу, називають шрифтом.

Існує багато різних репертуарів символів. Ця різноманітність пояснюється, передусім, особливостями алфавітів різних мов. Крім того, деяким користувачам для роботи потрібні спеціальні символи. Так, видавці використовують символ ©, гравцям у бридж потрібні карти різних мастей (♣, ♦, ♥, ♠), а математикам – спеціальні символи (Σ , $\sqrt{}$, $\int$).

Як кодуються набори символів форматів файлів в Інтернеті?

Більшість людей представляють символи двома способами: за допомогою певного зображення, наприклад, & або I, за допомогою певного імені. Наприклад, наведений вище символ & називається амперсандом. Багато форматів файлів, які застосовуються в Інтернеті, використовують імена для менш поширених символів. Наприклад, у мові HTML (гіпертекстова мова опису документів - HyperText Markup Language, який масово застосовується в Інтернеті) використовуються такі імена, як & для & і &lgrave; для I. Усі імена в HTML починаються амперсандом і закінчуються крапкою з комою.

Це не зовсім коректний підхід, оскільки імена, в свою чергу, це послідовність символів. Для комп'ютера ж деякі символи необхідно подати у вигляді чисел, з якими він маніпулює найприроднішим чином. При достатній кількості символів, поданих таким чином, їх можна використовувати для написання інших імен. Для цього існують два підходи. У першому варіанті використовується набір кодів символів (coded character set) і кожному символу присвоюється певне число. У другому варіанті – кодуванні символів (character encoding) – послідовність символів подається у вигляді послідовності значень байтів.

Набір кодів-символів передбачає, що кожний символ – це єдине число. Наприклад, набір кодів символів ISO Latin 1 для країн Західної Європи користується числом 65 для кодування символу A, 126 для ~ і 241 для ñ. Якщо маємо певну послідовність чисел, можна просто визначити, якому числу в таблиці відповідає даний символ.

Звичайно, в різних країнах і мовах користуються різними репертуарами символів. Найзручніші для кодування символів – числа від нуля до 255 (будь-які значення одного байта). Маючи в розпорядженні лише 256 чисел, неможливо дати унікальний код будь-якому символу. Тому були розроблені різні набори кодів символів. Вищезгаданий набір кодів символів ISO

Latin 1 створений організацією ISO (International Organization for Standardization, Міжнародна організація по стандартизації) для подання всіх символів, необхідних для певної групи мов (у даному разі для мов Західної Європи з латинським алфавітом). Деякі набори кодів намагаються охопити інші групи символів. Найпопулярніші комп'ютерні системи мають власні специфічні варіанти наборів символів (так, у MS-DOS і Windows використовується кодові сторінки фірми IBM).

Найпростіший спосіб кодування символів базується на єдиному наборі кодів символів, що містить 256 (або менше) кодів. Якщо в розпорядженні користувача є текстовий файл, який закодований таким способом, можна взяти з нього будь-який байт і визначити по таблиці, який символ відповідає цьому значенню байта. Коли один текстовий файл використовує відразу декілька наборів кодів символів, процедура дещо ускладнюється. У такому разі існують спеціальні коди символів, що інформують програму перегляду файла про те, коли вона має переключитися на інший набір кодів символів. В іншому міжнародному стандарті, ISO 2022, описаний один із способів переключення між наборами кодів символів.

Такі мови, як китайська, містять більше 256 символів, тому для кодування кожного з них використовують декілька байтів. Причому при такому кодуванні можливі різні підходи. В одному з варіантів кожному символу відповідає один байт, а для подання всього спектру символів використовується декілька наборів кодів. При іншому підході кожному символу відповідає декілька байтів. Для економії місця ці два підходи часто комбінуються: деякі символи кодуються за допомогою одного байта, тоді як для інших використовуються два чи більше байтів. Читати такі файли (наприклад, з текстом китайською мовою), звичайно, складніше від мов, які використовують підхід один байт – один символ, поширений в американських та європейських програмах.

Однією зі спроб узагальнення цих підходів є стандарт Unicode, відомий як ISO 10646. В Unicode для кодування символів використовується діапазон чисел від нуля до 65 536 (256×256). Такий широкий діапазон дозволяє подавати в числовому вигляді символи мов з будь-якого куточка планети. Багато міжнародних стандартів підтримки намагаються дотримуватися Unicode. Можливо, одна з наступних версій HTML також підтримуватиме цей стандарт. Нині для кодування символів мовою HTML як стандарт використовується набір кодів символів ISO Latin 1. Хоча здійснюються й серйозні спроби зробити так, щоб мова HTML підтримувала ширший репертуар символів.

З розглянутих системних позицій проблеми кодування різноманітних символів уявимо суть оформлення тексту в Інтернеті.

Інтернет – це вільне об'єднання комп'ютерів. Всесвітня мережа вважає об'єднати інтереси користувачів комп'ютерів з різними операційними системами: Macintosh, Unix, MS-DOS, Amigo, Atari, OS/2, Windows та ін. Внаслідок такої консолідації поступово зникають межі між комп'ютерами, побудованими на різних апаратурних і програмних платформах. Все це стає можливим, якщо немає плутанини з різними форматами файлів, якими обмінюються користувачі мережі. Чим ширший спектр комп'ютерних систем, на яких можна працювати з певним форматом файлів користувача, тим більше у нього шансів стати популярним в Інтернеті.

Інтернет – мережа міжнародна. Багато років Інтернет переважно був мережею Сполучених Штатів, і тому американська англійська негласно була офіційною мовою Інтернету. Нині ситуація суттєво змінюється. Навіть, коли текстові файли написані американською англійською та іншими мовами, часто використовують не ISO Latin 1, а іншу кодову таблицю (ASCII, KOI8, Win 1251 та ін.).

На це слід зважати, оскільки навіть у Сполучених Штатах кодування символів відбувається по-різному на різних комп'ютерних системах.

Оскільки різні комп'ютерні системи використовують різні набори кодів символів, ця проблема стала загальнопоширеною. Її можна вирішити двома способами. Перший передбачає використання усіма однієї системи кодування, наприклад стандарту Unicode (що, очевидно, станеться не скоро). Другий – в кожному текстовому повідомленні необхідно чітко зазначати, яке кодування (кодова таблиця) символів використовується. Інтелектуальне програмне забезпечення відповідного комп'ютера зможе коректно відтворити це повідомлення. З появою нових стандартів програмного забезпечення друга можливість стає реальнішою.

Текстові файли передаються у вигляді простого тексту (plain text). На жаль, простий у даному разі не є синонімом слова нескладний. Простий текст важко зробити таким, щоб він легко читався і привертав увагу, оскільки в ньому немає шрифтів різних написань, графіки, заголовків, підзаголовків, приміток тощо. Ці додаткові особливості мають назву розмітки тексту (markup). Один із найпростіших способів розмітки тексту передбачає використання імен для позначення спеціальних символів. Нижче розглянемо інші способи оформлення тексту, зокрема, фізичну та логічну розмітку тексту.

Насамперед слід усвідомити різницю між фізичною (physical) і логічною (logical) розмітками тексту. При використанні фізичної розмітки зазначається точний вигляд кожного фрагмента тексту, наприклад, центрований текст, 14-м кеглем, жирний, стиснутий, гарнітура Futura. При логічній розмітці тексту – логічне значення даного фрагмента, наприклад, це заголо-

вок розділу.

Ці два способи розмітки використовуються в різних ситуаціях. Щоб надрукувати текст на принтері, необхідно користуватися фізичною розміткою, тобто слід визначити розмір полів, формат приміток, величину абзацного відступу на початку кожного параграфа. Деякі версії текстових процесорів застосовували лише фізичний тип розмітки. При цьому для кожного фрагменту тексту необхідно зазначати його шрифт, розмір і стиль.

При обміні інформацією з іншими користувачами до фізичного оформлення тексту висувають деякі обмеження. Наприклад, стандартний аркуш паперу в різних країнах має різний розмір. І те, що має непоганий вигляд на папері, який відповідає американським стандартам, виглядатиме не досить добре на довгому і вузькому аркуші формату А4, який використовується в Європі. Із електронними документами, наприклад, з документами, отриманими з Інтернету, ситуація ще гірша. Розмір екрана, його розділення, шрифти і підтримка графіки – всі ці параметри змінюються залежно від комп'ютерної системи. Тому краще зробити відповідність документа існуючому дисплею, знову зробити форматування його.

Частіше комп'ютерні додатки вживають логічне оформлення тексту, для якого кожна частина документа пов'язується з його логічним значенням. Наприклад, деякі слова матимуть позначку виділити, іноді курсив. При друкуванні документа або виведенні його на екран логічний формат конвертується у відповідний до даної ситуації фізичний формат. Слова, які необхідно виділити на системах, що не підтримують курсив, будуть підкреслені. У тих країнах, де прийнято напівжирний шрифт, позначені слова виділятимуться за його допомогою.

У деяких ситуаціях логічне оформлення дуже важливе, наприклад, у разі обміну електронними документами типу сторінок World Wide Web або при створенні та публікації великих праць (таких, як книги). Багато видавців зберігають книги в електронній формі, використовуючи SGML (Standard Generic Markup Language, стандартна узагальнена мова розмітки документів). Такий підхід допомагає спростити процес створення книги (не слід постійно стежити за тим, щоб зберігалися розмітка сторінки і вибрані шрифти), при цьому можливо легко змінювати розмір книги та її формат.

Перетворення логічної розмітки у фізичну здійснюється за допомогою таблиці стилів (style sheet), де наводяться способи відображення кожного логічного елемента. Пізні системи керування процесом перетворення може відрізнятися деталями. В одних випадках логічна розмітка задається за допомогою текстових команд, і вхідний документ обробляється для створення вихідного документа, що містить фізичну розмітку. В інших – для логічного подання використовується подвійний формат текстового процесора,

а користувач редагує остаточний варіант фізично розміченого документа.

На завершальній стадії підготовки документа до використання його в Інтернеті треба зберегти розмітку. Це здійснюється таким чином. Для передачі даних на інший комп'ютер доцільно подати їх у текстовому вигляді. Якщо ж треба зберегти розмітку, можливий один із трьох підходів.

Перший підхід передбачає можливість зберегти розмітку, включивши її опис до тексту, наприклад, подавши фрагмент тексту: ...Прийняти **правильне** рішення... у вигляді ...Прийняти <bold> правильне <end bold> рішення.... Перевага такого підходу полягає в тому, що файл залишається текстовим (хоча, звичайно, виглядає він трохи незвично), а це значить, що його легко передавати з комп'ютера на комп'ютер. Якщо в користувача комп'ютера є програмне забезпечення, що підтримує цей формат, можна відновити документ у початковому вигляді. Існує чимало способів подання розмітки тексту, один із яких – HyperText Markup Language (HTML), що використовується в World Wide Web.

За допомогою другого підходу зберегти розмітку можна передаючи зображення кожної сторінки. Так працюють факс-машини, вони створюють графічне зображення кожної сторінки і передають його. Недоліком є те, що одержувач у результаті має лише зображення сторінки. Тому, щоб відредагувати документ, одержувач повинен ще раз його набрати. Проте цей недолік стає перевагою в тому разі, коли потрібно дати одержувачу можливість внести зміни до відправленого документа. Факс-модем дозволяє передавати документи з одного комп'ютера на інший. Незважаючи на те, що переданий документ має текстовий формат, його досить важко перетворити так, щоб вміст файла можна було відредагувати. Доцільно набрати все спочатку.

Третій підхід збереження розмітки реалізується через розробку особливих файлів, що містять не лише текст, а й інформацію про особливості його оформлення. Більшість текстових процесорів і настільних видавничих систем використовують такий підхід. Головна проблема полягає в тому, що в результаті майже всі текстові процесори та настільні видавничі системи застосують свої формати. Найдорожчі програми можуть читати файли, створені за допомогою програм-конкурентів. А тому слід використовувати спеціальні формати файлів для створення загальнодоступних документів.

В Інтернеті зберігання розмітки здійснюється за допомогою першого підходу із застосуванням не лише мови HyperText Markup Language (HTML), але й її модифікацію - мови Dynamic HyperText Markup Language (DHTML), яка суттєво розширює можливості Інтернету завдяки створенню інтерактивних Web-сторінок та нових стандартів зв'язку з базами даних.

Ми розглянули основні варіанти вирішення проблеми кодування інфор-

мації в Інтернеті з урахуванням можливих наборів символів у різноманітних моделях комп'ютерів та їх програмного забезпечення. Це дозволить опанувати механізм здійснення процедури доступу до інформаційних ресурсів Інтернету при використанні різномовних документів та відповідного програмного забезпечення для різних комп'ютерних платформ, підвищити ефективність використання глобальної мережі в документно-комунікаційній сфері.

СПИСОК ЛІТЕРАТУРИ

1. Кент П. *Использование Netscape Communicator 4: Пер. с англ.* – К.: Дialeктика, 1997. – 336 с.: ил.
2. Кендл Т. *Форматы файлов в Internet: Пер. с англ.* – СПб: Питер, 1997. – 320 с.: ил.
3. Смоляницкий Б.М. *Информационное обслуживание в Интернете. Ч. II. Учеб. пособие для студентов ХГИК по спец. Документоведение и информ. деятельность.* – Х.: ХГИК, 1998. – 64 с.
4. *Інформаційний сервіс в Інтернеті: Навч. посіб.* / В. М. Шейко, Б. М. Смоляницький, Л. Я. Філіппова та ін.; За ред. В. М. Шейка. – Х.: ХДАК, 1998. – 208 с.

Надійшла до редколегії 21.04.2000

УДК 378.6

І.О. ДАВИДОВА

УПРАВЛІННЯ ІНФОРМАЦІЄЮ – ІННОВАЦІЙНІСТЬ ФАХОВОЇ СВІДОМОСТІ

Розглядаються питання викладання курсу Інформаційний менеджмент і маркетинг, завдання, складові частини курсу, міжпредметні зв'язки

Життя стверджує прогноз Національного наукового фонду США, виконаний Центром кількісного аналізу наук понад 20 років тому. Розглядаючи перспективи розвитку наукової комунікації до 2000 р., його автори визначили 3 головні чинники розвитку: суспільний, технологічний, організаційний. Останній передбачає підвищення ролі інформації в удосконаленні управління, у процесах прийняття рішень і в досягненні економічної ефективності функціонування підприємств. Саме сфера управління інформацією є одним із напрямів, у якому очікується подальший прогрес наукової комунікації. Практика свідчить, що в організаціях уже нині виникла потреба у фахівцях, які мають навички професійного управління процесами інформаци-