

6. TVORBA TEXTOVÝCH DOKUMENTOV

V tejto kapitole sa stručne zoznámime s dejinami písma, jeho šírenia a archivácie, so základmi typografie, s tvorbou a úpravou textových dokumentov pomocou IKT.

Text (z Latinčiny *textus*) je záznam, vizuálna podoba hovorenej reči, môže byť ručne napísaný, vytlačený na papier alebo iné médium, vytesaný do skaly, vtlačený do hlíny, vygravírovaný v kove alebo zobrazený na monitore počítača, bankomatu či mobilného telefónu. Čítanie napísaného textu je náhrada orálnej prezentácie, ktorá bola pôvodným spôsobom komunikácie medzi homo sapiens. Existuje veľa definícií slova text. Z pohľadu informatiky je to je postupnosť znakov reprezentovaných binárnym kódom. Obyčajne jeden znak tvorí päť až osem bitov - jeden *байт* (*byte*).

6.1 STRUČNÉ DEJINY A FORMY PÍSMO

Písmo sa vyvíjalo súčasne a nezávisle na niekoľkých miestach na svete - Číne, Indii - Harapa v súčasnosti Pakistan, Mezopotámii na území dnešného Iraku - sumerské klinové písmo, Egypte, Amerike - Peru, Mexiku. Naše písmo *Latinka* má korene na Strednom Východe - Judei, Fenícii, Grécku a Ríme. Má históriu dlhú skoro 4000 rokov. Vychádza zo spoločného základu *proto kanájskej* - *proto sinajskej* abecedy, ktorá adaptovala niektoré *egyptské fonetické hieroglyfy* pre samostatné spoluhlásky. *Abeceda*, v Gréčtine *alfabet* podľa prvého písmena alfa je adaptáciou písmena alef ktorého tvar sa vyvinul zo zvuku alef, ktorý označoval vola. Podobne beta pochádza od beth čo znamená dom, znie ako b píše sa ako znak domu. Z tohto písma sa vyvinulo *ugaritické*, *juho-arabské* a *fénické* písmo. Z neho sa vyvinulo staro *hebrejské*, *aramejské* a *archaické grécke* písmo. Z aramejského písma vzniklo niekoľko písiev medzi nimi *hebrejské*, *arabské* a *perzské písmo*. Z archaického gréckeho písma sa vyvinulo *grécke písmo* a jeho deriváty ako *Hlaholika* a *Cyrilika* ako aj *etruské* a *latinské písmo (rímske)*, teda to ktoré používame v prevažnej časti Európy. Číslice, ktoré používame majú *indo-arabský* pôvod.

V súčasnosti sa používa v Európe, Amerike, Austrálii a mnohých ázijských a afrických krajinách *Latinka*. Grécke písmo a jeho odvodeniny sa používajú v Grécku, Bulharsku, bývalých sovietskych republikách (aj v Strednej Ázii) a bývalej Juhoslávii (okrem Slovinska a Chorvátska). V Ázii sa používajú fonetické abecedy - *arabské* a *hebrejské písmo* v krajinách Blízkeho Východu aj niektorých afrických zemiach, *gruzínske písmo*, písma niekoľko abecied v Indii a iných krajinách ako aj znakové písma v Číne, Kórei a Japonsku. Niektoré texty - *arabské* a *hebrejské* sa píše opačne ako u nás, to je od prava do ľava.

Okrem týchto klasických druhov písma vznikli špeciálne písma napríklad *slepecké písmo* skladajúce sa z bodiek usporiadaných v matici, *Morzeova abeceda* skladajúca sa z bodiek, čiarok a medzier, *čiarové písmo (bar code)* ktoré predstavujú inú vizuálnu alebo zvukovú reprezentáciu tej istej abecedy, ktorú používame.

Záznam textu

Texty sa zaznamenávali a dodnes sa zaznamenávajú ručne alebo strojovo na *kameň* kreslením, maľovaním, rytím - Altamíra, Mezopotámia, Perzia, Egypt, v súčasnosti gravírovacie stroje riadené počítačom, *hlíny* - Mezopotámia, Harapa, na papyrus - Egypt, pergamon - Grécko, na

ÚVOD DO INFORMATIKY

(c) 2008 Doc.Ing. Martin Šperka, PhD.

(Materiál neprešiel odbornou ani jazykovou korekciou a je určený pre študentov Fakulty informatiky Bratislavskej vysokej školy práva.)

hodváb a papier - Čína, ako uzlíky na špagáte - Inkovia. Papier mal veľký význam v rozvoji písomníctva. Umožnil - spolu s pergamenom viazať popísané listy do kníh, čo v tej dobe znamenalo maximálne množstvo údajov na objem a váhu nosičov informácie. Knihy sa reprodukovali ručným opisovaním, hlavne v kláštoroch, ktoré predstavovali centrá vzdelanosti nielen v Európe ale aj v Indii, Tibete a Japonsku a kde vznikali knižnice, ktoré predstavujú ešte aj dnes najväčšie kolekcie dát. Revolúciu v archivovaní a šírení informácie pomocou textu aj s obrazom znamenal vynález tlače. Tlač pomocou kamenných a drevených štočkov vznikla v druhom storočí n.l. v Číne. Z Číny cez Perziu, Arabov a Osmanskú ríšu sa dostala do Európy. Masové šírenie textovej informácie prešlo dramatickým vývojom. V pätnástom storočí Johannes Gutenberg vytlačil v Nemecku prvú knihu v Európe. V kníhtlač narážala aj na odpor - jeden z argumentov proti bol, že aj tak väčšina ľudí nevie čítať. No už za 50 rokov sa vytlačilo až 6 miliónov kníh. V roku 1488 vznikla prvá tlačiareň v Anglicku – Westminster Abbey. V minulom storočí sa začali používať kovové štočky. Tento princíp používajú aj klasické písacie stroje ako aj stroje s guľovou hlavou - IBM patent používaný v elektrických písacích strojoch používaných ako terminál počítača. Podobne aj počítačové valcové tlačiarne. Ďalej sa tlač vyvíjala s technológiami ako umelohmotné štočky, ihličkové tlačiarne, bublinková tlač, laserové tlačiarne, termotransfer atď.



Obr. 6.1. Strana z Gutenbergovej biblie

Tlač v počítačovej ére

Tlač kníh, časopisov a novín sa neustále zdokonaľovala a mechanizovala ale najrýchlejší rozvoj tvorby tvorby, archivácie a prenosu textových dokumentov spôsobil vynález počítačov a počítačových sietí. História textových elektronických dokumentov ako aj komunikácie pomocou písaného a hovoreného textu by sme mohli stručne rozdeliť na štyri obdobia:

1. Sedemdesiate roky - sálové počítače, komplikované ovládanie, *textovo orientované programy*. Drahé tlačiarne s jedným typom a veľkosťou písma, prenos textu d'alekopisom.
2. Osemdesiate roky - mikropočítače a PC, WYSIWYG - programy typu Word processor (WordStar, Chi-writer, T602, MS Write). *Možnosť tlače textu a grafiky. Prenos písaných dokumentov elektronickou poštou alebo faxom. Počiatky rozpoznávania tlačeného textu.*
3. Deväťdesiate roky - rýchle PC, veľa pamäti, integrovanie textu a obrazu typu text processor (AmiPro, MS Word) a DTP (Desk Top Publishing - Page Maker, Ventura Publisher), *kontrola gramatiky, hypertext, on-line dokumenty - Internet. Kvalitná farebná tlač na*

domácich počítačoch. Elektronická pošta, internetové diskusné fóra založené na písaní textu, fax v domácnostiach. Rozpoznávanie písaného textu na mobilných zariadeniach.

4. Tretie tisícročie - sémantická pavučina, dolovanie vedomostí, automatický preklad, prepis textu z plynulej hovorenej reči, automatické komponovanie textu na základe neúplnej špecifikácie obsahu, a ešte nepredvídané možnosti. Napriek predpovediam o limitovaní používania tlačených dokumentov, vzhľadom na nízku cenu a kvalitu tlače nárast množstva tlačených dokumentov ale aj možnosti recyklovateľnej tlače na *elektronický a virtuálny papier, elektronické zápisníky integrované s telefónmi a bezdrôtovým prenosom dát.*

6.2 REPREZENTÁCIE ZNAKOV, ŠTRUKTÚRA TEXTU

Základom textu sú abecedné, číselné a špeciálne znaky. Základnou dátovou jednotkou v počítačoch je bajt, ktorý dovoľuje zakódovať 256 rôznych znakov. V podstate to postačuje, ale nakoľko vo svete existuje veľa typov abecedy, vznikajú tu problémy s kódovaním. Prvým úspešným štandardom pre kódovanie znakov bol *ASCII (American Standard Code for Information Exchange)*, ktorý špecifikoval fixne prvých 128 znakov sady a zvyšných 128 sa líšilo v závislosti od krajiny a kód. Ďalej je to *ISO (International Standards Organization) 8859-2* tiež *Latin-2* alebo *CP 852*. Iným univerzálnym štandardom je *ANSI 1250* známy ako *Windows 1250* a *Unicode*.

Na rozdiel od predchádzajúcich štandardov sa Unicode snaží zahrnúť znaky v každej známej abecede. Vďaka tomu možno v jednom textovom dokumente zapísať texty z rôznych jazykov používajúcich rôzne písma. Kódovanie prebieha v dvoch krokoch. V prvom sa znaku priradí nejaké prirodzené číslo. V druhom sa toto číslo zapíše ako postupnosť bajtov. Existuje viacero možných kódovaní, napríklad *UTF-8* alebo *UTF-16 (Unicode Transformation Format)*. Toto kódovanie navrhla nezisková organizácia s názvom *The Unicode Consortium*, ktorej cieľom bolo vytvoriť jednotnú sadu znakov *UCS (Universal Character Set)*. Unicode používa dvojbajtové kódy - to je 65 536 znakov. Štandard rozdeľuje znaky do sedemnástich dvojbajtových rovín čo umožňuje definovať až $1\,114\,112$ to je 17×2^{16} znakov. Tento štandard je implementovaný v moderných operačných systémoch a v jazykoch XML a JAVA.

Textové dokumenty tvoria skvencie znakov. Tieto sú nejakým spôsobom štruktúrované. Rozoznávame *logickú štruktúru a fyzický vzhľad textu*. Autor textu sa sústreďuje na obsah (sémantiku) a nezaobrá sa zalomením textu.

Jednoduchý text

Je neštruktúrovaná alebo len vizuálne štruktúrovaná postupnosť znakov - *reťazec (string)* znakov. Napríklad *TOTOJETEXTVKTOROMSAPOUZIVAABECEDA*. Text je zrozumiteľnejší ak slová rozdelíme medzerami, čiarkami a bodkami. Reťazec *ABECEDA* za ktorým je medzera v *ASCII* kóde reprezentovanom hexadecimálnom kódom je postupnosť *41 42 45 43 45 44 41 20*. Po prevode na binárny kód a zápisom tak, že každý bajt znázorníme v zvláštnom riadku, ktorý reprezentuje adresu v pamäti dostaneme obraz pamäti počítača s uvedeným textovým reťazcom. Text na najnižšej úrovni je reťazec jednotiek a núl, pričom jednotlivé znaky reprezentujú bajty – ak jeden znak reprezentuje jeden bajt, alebo dvojice bajtov.

adresa	0100 0001	A
adresa + 1	0100 0010	B

ÚVOD DO INFORMATIKY

(c) 2008 Doc.Ing. Martin Šperka, PhD.

(Materiál neprešiel odbornou ani jazykovou korekciou a je určený pre študentov Fakulty informatiky Bratislavskej vysokej školy práva.)

adresa + 2	0100 0100	E
adresa + 3	0100 0011	C
adresa + 4	0100 0100	E
adresa + 5	0100 0101	D
adresa + 6	0100 0001	A
adresa + 7	0010 0000	medzera

Značkovaný text (marked-up text)

Značkovaný text je jednoduchý text obohatený o značky, ktoré môžu byť ľubovoľné znaky vrátane znakov abecedy. Značka začína bodkou. Príkladom programu pre prácu so značkoványm textom je Troff v operačnom systéme UNIX

.ps 20	veľkosť v bodoch
.vs 22	vertikálna medzera
.ft B	typ písma
.ce	center
.sp	medzera

Iným príkladom je *TEX* - Donald. Knuth v roku 1984 a *LATEX* - Lamport v roku 1986. Cieľom je vysokokvalitná sadzba hlavne matematických článkov. Príklad v AMS-LaTeX. V tomto prípade značka začína opačným lomítkom :

```
\title{Názov kapitoly}
\author{Michal Písmo}
\date{16 jun 2007}
\documentstyle{document}
\maketitle
\[x^{2n} x^{y^z} y_m\]
\[y= \sum_{i=1}^n x_i\]
\[x=\int_0^1 f\]
```

LaTeX je pre technické články, AMS-TeX - pre matematické články, mathematical, BibTeX - pre bibliografické data bázy, SliTeX - pre prezentácie (slajdy)

Potenciálnym problémom značkovaného textu je miešanie logickej štruktúry so *zalomením (layout)* textu. Normou ktorá toto rieši je ISO 8879:1986 - jazyk *SGML (Standard Generalized Markup Language)* ktorý oddeľuje vzhľad od logickej štruktúry textu použitím typov dokumentov.

K značkovacím textom patrí aj *STML Spoken Text Markup Language* alebo *JSML Java Speech Markup Language*. Sú to jazyky pre popis textu, ktorý sa *syntetizuje do zvukovej podoby (text to speech)*, napríklad v internetovom prehliadači. Podobne existujú aj značkovacie jazyky pre popis hudby.

Štrukturovaný text

Keď sa používajú *značky (tags)* je štruktúrna informácia rozdelená po celom dokumente. Aby sme túto štruktúru extrahovali a potom efektívne spracovali, editor dokumentu musí prehľadávať a analyzovať celý text.

Štrukturovaný text používa pre logickú a vizuálnu organizáciu dátové štruktúry. Napríklad kapitoly alebo odstavce s názvami. Obyčajne je to stromová štruktúra. Podobne vzhľad môže mať stromovú štruktúru - strany, bloky, stĺpce. Príkladom je *ODA (Office Document Architecture)* - štandard ISO z roku 1989 pre štruktúru dokumentov.

Hypertext

Tradične je textový dokument - správa, poviedka, román lineárny text, ktorý sa číta od začiatku do konca aj keď samotný dej môže byť nelineárny to znamená že tok času poviedky nemusí byť zhodný s časovou osou textu dokumentu. Výkladový slovník je typickým príkladom klasickej realizácie nelineárneho textu, kde sa postup čítania vetví podľa odkazov na pojmy, ktorými sa vysvetľuje hľadaný pojem. Proces čítania elektronických nelineárnych textových dokumentov a zprístupňovania dát sa v súčasnosti neobmedzuje obmedzený na počítač pri ktorom sedíme ale stále viac sa prenáša aj na iné počítače - v tej istej budove, meste, krajine, svetadiely alebo zemeguli.

Hypertext môže stručne charakterizovať ako text v ktorom sú odkazy na iné časti textu a ktoré sa rýchlo sprístupnia ukázaním na odkaz. Tento sa vizuálne charakterizuje zvýraznením - tučným písmom, kurzívou, podčiarknutím alebo farbou. Mechanizmus navigácie v dátach pomocou hypertextu pripomína štruktúru deja rozprávky L. Carolla "Alica v ríši zázrakov", kde sa malá hrdinka vnára do rôznych prostredí a dejov a potom sa opäť vracia tam, odkiaľ prišla. *Hypertext predstavuje aj spôsob reprezentácie vedomostí, kde nové pojmy sa vysvetľujú známymi pojmami.* Budovanie vedomostí v tomto prípade predstavuje hierarchickú štruktúru pojmov. Prvým programom umožňujúcim tvorbu hypertextových dokumentov bol *Hypercard* na počítačoch *Apple MacIntosh*. Prvým hypertextovým systémom v internete bol systém *Gopher* - predchodca *WWW*.

Čítanie nelineárneho textu je teda veľmi pohodlné, obzvlášť v takých prípadoch ak potrebujeme vedieť definície alebo bližšie špecifikácie pojmov ktoré sa v texte používajú. Hypertext však predstavuje aj nebezpečenstvo, že čitateľ v texte "zablúdi". Preto je nevyhnutné aby prehliadač takéhoto textu mal možnosť návratu na najbližšiu vyššiu úroveň, respektíve na určitú, predom označenú úroveň. Pred tvorbou hypertextového dokumentu je treba pozorne zvažovať jeho štrukturovanie.

V súčasnosti existuje veľa programových produktov umožňujúcich prácu s hypertextom. Aj mnohé programy pre spracovanie textu umožňujú import a export hypertextových dokumentov. Normou ISO a IEC (International Electrotechnical Committee) v oblasti hypermédií je jazyk *MHEG (Multimedia/hypertext Expert Group)*. Jedným z najrozšírenejším štandardom je podmnožina *SGML* - jazyk *HTML (Hyper Text Markup Language)*, ktorý je základom *WWW* dokumentov. Tento podlieha neustálym zmenám a o jeho rozvoj sa stará *W3C (World Wide Web Consortium)*.

Metadáta

Metadáta sú také dáta (textové), ktoré popisujú iné dáta, špecifikujú čo iné dáta znamenajú. Častokrát môžeme z kontextu dedukovať čo daný text opisuje, ale aby to bolo jednoznačné musíme ho označiť. Napríklad ak máme adresár osôb tak každú položku napríklad meno alebo ulicu označíme značkou, ktorá špecifikuje tieto položky ako meno a ulica. V klasických databázových systémoch sú jednotlivé položky na presne vymedzených miestach v zázname. Pri použití značiek s metadátami môžu byť na ľubovoľnom mieste alebo poradí. Pre popis dát slúži jazyk popisu dát. V súčasnosti najrozšírenejším štandardom je podmnožina jazyka *SGML* - *XML (Extended Markup Language)*. Je to párový značkovací jazyk, kde je objekt

označený na začiatku a na konci značkou v špicatých zátvorkách. Na rozdiel od HTML kde sú len niektoré značky párové. Koncová značka začína lomítkom. Dvojice značiek sa môžu do seba vnárať. Napríklad záznam s menou a ulicou vyzerá nasledovne.

```
<Záznam><Meno> Ján Múdry</Meno><Ulica> Dolná 34 </Ulica></Záznam>
```

6.3 POČÍTAČOVÉ SPRACOVANIE TEXTU

Pri spracovaní textu pomocou počítača sa používajú rôzne operácie. Nasledujúci text stručne sumarizuje najviac používané.

Práca so znakmi a reťazcami (string processing)

Dobre spracovaná časť informatiky hlavne v oblasti prekladačov je lexikálna analýza a vetný rozbor - syntaktická analýza. Pre prácu s textovými reťazcami existujú rôzne nástroje v programovacích jazykoch napríklad spájanie hľadanie znaku v reťazci, spájanie reťazcov a podobne. Jeden z prvých jazykov určených pre spracovanie reťazcov bol Snobol. Jazyk C obsahuje množstvo funkcií pre prácu s reťazcami, podobne aj Java alebo Javascript ako aj aplikačné programy napríklad MS Excel.

Zostavovanie textu (Editing)

V Slovenčine sa používajú aj výrazy *redakcia* textu alebo *editovanie*. Existuje viac programov pre takúto aplikáciu. Najznámejší je *MS Office Word* alebo *Open Office Write*. V týchto programoch sa používajú funkcie ako *vkladanie*, *označenie*, *vyrezanie*, *zrušenie*, *kopírovanie reťazca*, *hľadanie*, *nahradenie reťazca*. Označovanie sa robí pomocou myši. Tieto programy majú veľké množstvo príkazov a bližšie sa nimi bude zaoberať špeciálna kapitola.

Formátovanie:

Môže byť *interaktívne s využitím princípu WYSIWIG* - What You See Is What You Get a *neinteraktívne* - na základe nejakého algoritmu. WISIWIG sa používa aj v programe Word a Write. Používateľ má okamžitú spätnú väzbu, vie ako bude text vyzerat' na papieri. S použitím značkovaného textu to nie je možné, napriek tomu mnoho ľudí používa napríklad LATEX tak, že rovnice píše pomocou abecedných a špeciálnych znakov. Výstupom formátovaného textu môže byť *bitová mapa* to je rastrový obraz alebo popis v *jazyku popisu dokumentu (Page Description Language)* napríklad *PostScript* alebo *Encapsulated PostScript* - formát *EPS*, zavedený firmou Adobe. Následníkom EPS je *PDF - Page Description Format*, ktorý zahrňuje aj hypertext.

Kompresia a šifrovanie textu

Pri archivovaní aj prenose textu sa získa veľa priestoru v pamätiach ako aj času pri prenose kompresiou. Kompresia využíva redundanciu informácie v správe. Existuje niekoľko kompresných metód textu. Niektoré využívajú fakt, že niektoré znaky sa vyskytujú v jazyku s vyššou pravdepodobnosťou, napríklad e a tieto kódujú kratším dvojkovým kódom ako tie čo sa vyskytujú menej, napríklad x. Toto používa Huffmanova metóda. Iné združujú postupnosti znakov do nových znakov napríklad metóda LZV používaná vo formátoch ZIP alebo ARJ.

V poslednej dobe v súvislosti s počítačovou bezpečnosťou získava na dôležitosť šifrovanie textovej informácie. Jednou metódou National Bureau of Standards z roku 1977 je *DES - Data Encryption Standard*. Flexibilnejšie a spoľahlivejšie algoritmy používajú algoritmus verejného kľúča .

Jazykovo špecifické operácie

Sú to napríklad programy pre rozdeľovanie riadkov podľa pravidiel jazyka, hľadanie chýb - *správne písanie slov (spell checker)*, gramatické pravidlá, vysvetlenie významu zvoleného slova a ponuka synonym podľa významového slovníku, počítanie počtu znakov a/alebo slov, štatistická analýza.

6.4 VZHĽAD TEXTOVÉHO DOKUMENTU

Dôležitým vizuálnym atribútom je *vzhľad textového dokumentu* - rozloženie jednotlivých častí textu, obrázkov (*layout*).

Typografia písma

Typografia je náuka o písme, ale aj remeslo a umenie - napríklad kaligrafia a písomného prejavu. Hlavné zásady typografie by mal poznať každý kto chce publikovať v papierovej alebo elektronickej forme. Hlavným kritériom je *čitateľnosť*, to je schopnosť nájsť, identifikovať, rozlíšiť a absorbovať text. Čitateľnosť sa podstatne zvýši pozorným výberom typu a rozložením písmen. Písmo je charakterizované druhom - *rezom (typeface)* písma. Existujú dve hlavné skupiny *pätkové (serif)* a *(bezpätkové - sans serif)* písmo. Príkladom prvej skupiny je *Times Roman* a druhej *Helvetica*. Sadzba písma môže byť normálna alebo zvýraznená. K zvýrazneniu sa používa *kurzíva (Italic)*, **tučné (fat)** alebo podčiarknuté (underlined) písmo alebo p r e l o ž e n á sadzba. Písmená môžu mať rôznu veľkosť, v klasickej tlači to bolo pár veľkostí nakoľko písmená sa sádzali ručne a bolo by treba veľa veľkostí štočkov. Veľkosti mali názvy ako petit, garmond a cicero. V počítačoch sa veľkosť môže nastaviť ľubovoľne a udáva sa v bodoch, metrických alebo anglosaských dĺžkových jednotkách – cm, inch. Používajú sa rôzne medzery medzi písmenami, slovami, riadkami. Vzhľad textového dokumentu ďalej určuje *počet a tvar stĺpcov* - bloky textu s obdĺžnikovým alebo iným tvarom, ich vzdialenosťami od okrajov ako aj navzájom medzi sebou, farbou písmen a pozadia.

Veľký prínos k typografii znamenala abeceda Francúza *C. Garamnoda* z roku 1530. V roku 1757 navrhol *J. Baskerville* abecedu s extrémne veľkými pätkami. Jeho písmo sa v takmer nezmenenej podobe používa dodnes. V roku 1816 *W. Calson* navrhol medzi inými aj *písmo bez pätiiek*. Na tú dobu pôsobilo groteskne a je známe pod menom *Grotesk*. Prvé modely písacích strojov koncom minulého storočia používali písmo *G. Bodoniho* (1740-1813). Fotosadzba v polovici tohto storočia umožnila ďalší rozvoj tlače.

Prínos k počítačovej typografii urobil *Donald Knut*, autor niekoľkozvážkovej zbierky počítačových algoritmov. Namiesto bodového písma, ktoré sa používalo v počiatočných počítačových spracovaniach textu použil písmená vytvorené z kriviek -čiarová grafika, čím sa umožnilo plynulé zmenšovanie a zväčšovanie písmen bez úkmy na kvalite - písmená zložené z bodiek sa nedajú plynule zväčšovať.

Aj keď sa typy písmen ako aj rozloženie textu počas stáročí zmenili, platia určité základné princípy: premenlivé rozostupy písmen kde napríklad "l" je omnoho užšie ako "m". Adaptívne zmenšovanie rozostupov šetrí priestor a spôsobuje, že text je čitateľnejší ako pri konštantných rozostupoch. Čitateľné písmo je veľké 9 alebo 10 bodov (1 bod = 1/72 palce). Pätkové písmo má o niečo lepšiu čitateľnosť ako *bezpätkové*.

ÚVOD DO INFORMATIKY

(c) 2008 Doc.Ing. Martin Šperka, PhD.

(Materiál neprešiel odbornou ani jazykovou korekciou a je určený pre študentov Fakulty informatiky Bratislavskej vysokej školy práva.)

Dôležitou charakteristikou textových blokov je *zarovnanie* - doľava, na stred alebo doprava. Podľa niektorých prameňov zarovnaný text znižuje rýchlosť čítania o 12 percent (kvôli väčším medzerám je v riadku menej slov) . Veľké písmená zaberajú o 30 percent väčší priestor a znižujú rýchlosť čítania o 12 percent.

Všeobecným pravidlom je, že treba používať čo najmenej typov písma (1 až 3), veľkostí (1 až 3) a štýlov písma. Treba používať veľké (kapitálky, verzálky) aj malé písmená a vyhýbať sa používaniu len veľkých písmen. Dĺžka riadku by mala byť v rozmedzí 40 až 60 znakov. Rozloženie textu (Layout) má sledovať hrubý raster. Rozloženie môže byť symetrické alebo nesymetrické, ale nesmie sa miešať. Používajú sa 1,2 alebo 3 stĺpce. Medzery medzi nimi ako aj od okrajov strany majú byť primerané. Obrázky majú byť tiež zamerané v hlavnom rastrí.

Masový nástup osobných počítačov v osemdesiatich rokoch vôjde zrejme do histórie šírenia informácie pomocou písma tak ako aj posledná veľká udalosť - publikovanie v internete cez službu World Wide Web.

Logické členenie textu

Logická štruktúra sa prejaví aj vizuálne. Závisí od zamerania, obsahu a kontextu. Inú štruktúru má sms pomocou mobilného telefónu, elektronická pošta napríklad oznámenie o konferencii, diskusia - chat v internete, báseň, vedecký článok, román, učebnica, semestrálna alebo bakalárska práca. Podstata tvorby textu spočíva v generovaní myšlienok a ich usporiadaní do písomnej štruktúry vyhovujúcej potrebe študenta aj autora textu. Niektoré praktické zásady sú: Ciele textu je treba poriadne vymedziť. Treba používať krátke vety, zrozumiteľné, vyvarovať sa komplikovaným vetám a cudzím slovám - keď je potrebné uviesť ich v zátvorke. Text nesmie byť homogénny ale má mať vhodné členenie. Vety, súvetia, podkapitoly a kapitoly musia na seba nadväzovať a byť označené.

Predmet Propedeutika, v nasledujúcom semestri sa bude zaoberať spôsobmi písania rôznych dokumentov, potrebných pri štúdiu na vysokej škole.

6.5 TVORBA A SPRACOVANIE TEXTOVÝCH DOKUMENTOV

V tejto kapitole sa budeme zaoberať tvorbou, spracovaním, prenosom a archiváciou textových dokumentov pomocou IKT. Existuje veľa aplikačných programov, kde sa tvorí text. Najjednoduchšie aplikácie sú v bankomatoch, kde sa vyžaduje písanie jednoduchých textov, SMS a MMS, textové redakčné programy, zalamovacie programy a vkladanie textu v rôznych aplikáciách tvorby obrazových dokumentov.

SMS (Short Message Service) a MMS (Multimedia Message Service)

Predsťtuje najjednoduchšie textové dokumenty. Vznik tejto služby je náhodný, to znamená že nebol pôvodným zámerom operátorov mobilnej komunikácie. Jej úspech bol nečakaným prekvapením výrobcov mobilov a je príkladom revolučného rozšírenia technológie používateľmi zdola (*grassroot revolution*).

Pretože jednoduché mobilné telefóny nemajú klávesy všetkých znakov využíva sa číslícová klávesnica s niekoľkonásobným stlačením, čo spôsobuje ľuďom nezvyknutým na takýto spôsob písania problémy. Ale mladí ľudia sú schopní naučiť sa písať veľmi rýchlo. Aby sa komunikácia urýchlila používajú sa mnohé skratky napríklad "CuL8er" je skratkou "see you later". Pre vyjadrenie emócií sa používajú ikony výrazov ľudskej tváre (smiley) podobne ako

v elektronickej pošte. Rozšírením SMS je MMS kde správa môže obsahovať aj obrázok, zvukový a video záznam.

Textové redakčné programy (Text Editors)

Prvé textové redakčné programy sa používali pri písaní programov, mali len jeden typ a veľkosť písma. Keď sa v sedemdesiatich rokoch minulého storočia rozšírili polovodičové pamäti a tým aj počítačová grafika mohli vzniknúť programy, kde text na obrazovke mal tvar tlačeného textu *WYSIWYG (What You See Is What You Get)*. S šírením osobných počítačov v osemdesiatich rokoch začali vznikať programy ako WordStar, WordPerfect, Chi Writer, AmiPro. V bývalom Československu bol obľúbeným program T602. Tieto umožnili vytvoriť nový textový súbor (*new*), nahrat' súbor z disku alebo pásky (*load*), zapamätať súbor (*save*), vkladať (*insert*) text, rušiť (*delete*) znaky a slová, vystrihovať (*cut*), lepiť (*cut*), premiestňovať (*move*) zvolené reťazce znakov v rámci celého dokumentu a aj mimo neho, hľadať a nahrádzať inými reťazcami zvolený reťazec (*search and replace*), automaticky počítať počet znakov a slov. Neskôršie sa vkladali funkcie ako nastavenie veľkosti dokumentu, nastaviť veľkosť, typ a farbu znakov a pozadia, automatické zarovnanie textu, rozdeľovanie slov podľa gramatických pravidiel, korekcia písania (*spell checker*), slovník alternatívnych výrazov (*Thesaurus*).

MS Office Word a Open Office Writer

Súčasťou oknových operačných systémoch Microsoft Windows je textový editor *Notepad* (zápisník) a *Wordpad*. súčasťou balíka *MS Office*. Prvý z nich je jednoduchý textový editor, postačujúci pre mnohé aplikácie. Ten druhý je už dokonalejší a umožňuje mnoho funkcií, vymenovaných v predchádzajúcom odstavci. Program *MS Word* je pravdepodobne najrozšírenejším programom pre redakciu textu na svete. Jeho predchodcom bol program *MS Write*. Používa súbory typu .doc a rtf (RichText Format), ale umožňuje *import a export* do súborov typu .txt, html, pdf.

Program *Open Office Writer* je voľne dostupný program a má veľmi podobný vzhľad používateľského rozhrania a funkcie. Používa typy súborov .odt, .ott a .sxw ale poskytuje aj import a export súborov .doc. Export do .pdf je súčasťou programu a nepotrebuje licenciú.

Oba majú horeuvedené funkcie plus možnosti vkladať do textu rôzne objekty ako sú obrázky, hypertextové odkazy, video sekvencie, tabuľky z tabuľkového kalkulátora. Okrem toho je možné tvoriť aj vlastné objekty ako sú tabuľky, matematické vzorce a grafiky. Ďalej umožňuje nastaviť okraje textu, medzery medzi riadkami, odsadzovanie nadpisov, voliť počet stĺpcov textu.

Tvorba matematických textov

Používanie matematických textov - vzorce, rovnice je súčasťou štúdia na informatike. *MS Word* a *Open Office Math* poskytujú možnosti písať rôzne matematické symboly a písmená gréckej a hebrejskej abecedy, ktoré sa v matematike používajú. Ďalej sú to zlomky, hoirné a dolné indexy, mocniny, odmocniny, matice, sumy, integrály. Nasleduje ukážka tvorby objektov funkcie *Microsoft Equation* v príkaze *Insert* programu *MS Word*.

≤ (>) ⊗ ≥ ≠ ≡ ∞ ⇔ ∃ ∀ ¬ ∈ ∪ ∩ ⊆ ∫ ∫ ∇ ∞ ∑ ∏ α β γ δ ξ φ θ λ Ω Ψ

ÚVOD DO INFORMATIKY

(c) 2008 Doc.Ing. Martin Šperka, PhD.

(Materiál neprešiel odbornou ani jazykovou korekciou a je určený pre študentov Fakulty informatiky Bratislavskej vysokej školy práva.)

$$|x_1, y_1, z_1, w_1| \begin{pmatrix} abcd \\ efgh \\ ijkl \\ mnop \end{pmatrix} \sqrt[3]{a+b} \sum_{i=0}^5 2i \int_0^{100} idi \prod_{i=1}^{10} i \prod_{i=0}^{10} U_{ki}$$

Zalamovacie programy

Súčasné textové editory umožňujú tvorbu textu aj spolu s vkladaním obrázkov a tlačíť ich jednotlivo na laserových alebo atramentových tlačiarňach, alebo odoslať súbor na tlačiarenské stroje a tlačíť priamo knihy či časopisy. Iným spôsobom je tlač na priehľadné alebo matné fólie a z nich fotografickou cestou pripraviť podklady pre hromadnú tlač. Pre kvalitnú prípravu podkladov pre tlač sa používajú *zalamovacie programy* (*Desk Top Publishing - DTP*). Prvým takýmto programom bol program Ventura, ktorý pracoval aj pod systémom DOS. Neskôršie vznikli nástroje ako Adobe Page Maker a Quark Xpress. Tieto pracujú tak, že text sa importuje z iných programov (je možno tvoriť ho aj priamo v programe pre DTP) a text s obrázkami sa zalamuje. Funkcie sú podobné ako má Word ale je ich viac a pohodlnejšie sa pracuje hlavne s veľkými dokumentami. Príprava knihy s 500 stranami vo Word je náročný proces v porovnaní s použitím programov pre DTP.

ZDROJE INFORMÁCIÍ (kap 6)

ÚLOHY A MINIMÁLNE POŽIADAVKY (kap. 6)

ÚLOHY

Prečítajte si túto kapitolu a vyhľadajte v internete pojmy, s ktorými sa chcete bližšie oboznámiť.

Nájdite v internete tabuľky kódovania znakov ASCII, rozšíreného ASCII, ISO a Unicode a porovnajte ich.

Nainštalujte si voľne dostupný program Open Office a porovnajte výhody a nevýhody práce s programom MS Word.

V programe MS Word alebo OO Writer napíšte nejaký text a porovnávajte jeho vzhľad s rôznymi typmi a veľkosťami písma.

Pomocou niektorého vyhľadávača si nájdite stránku, kde je možné tvoriť rovnice pomocou jazyka LATEX, stiahnite si rovnice v obrazovej forme a vložte ich do textu.

Nájdite si nejaký súbor .html alebo .xml a pokúste sa nájsť súvislosť medzi textom dokumentu a jeho vzhľadom v prehliadači. Jazykom HTML sa budeme zaoberať v zvláštnej kapitole o tvorbe internetových dokumentov.

MINIMÁLNE POŽIADAVKY NA VEDOMOSTI

Vedieť stručné dejiny písma, počítačového spracovania textu, aké sú možné operácie s textom, aké sú hlavné zásady písania pre dobrú čitateľnosť?

Vysvetliť pojmy kódovanie znakov, ASCII, ISO, Unicode, štruktúrovaný text, značkovaný text, hypertext, typografia, veľkosť, typ písma, sadzba, pätkové, bezpätkové písmo.

MINIMÁLNE POŽIADAVKY NA ZRUČNOSTI

(Prečítajte si tieto činnosti bod po bode a ak niektoré z nich neovládáte, konzultujte ich s pedagógom)

ÚVOD DO INFORMATIKY

(c) 2008 Doc.Ing. Martin Šperka, PhD.

(Materiál neprešiel odbornou ani jazykovou korekciou a je určený pre študentov Fakulty informatiky Bratislavskej vysokej školy práva.)

Našartovať program Word, otvoriť nový dokument - súbor, napísať text s použitím rôznych typov, veľkostí a farieb písma, rôzne zarovnaním, s jedným alebo viac stĺpcami. Pri tvorbe textu vedieť používať označovanie, rušenie, kopírovanie a presun blokov textu. Text zapamätať na disk do súboru s určeným menom, exportovať do rôznych formátov súborov (.txt, .rtf, .pdf, .html, .ppt), načítať alebo premenovať súbor s príponou .doc, .rtf. Vedieť vytvoriť tabuľku s určeným počtom riadkov a stĺpcov, vložiť tabuľku načítaním z programu MS Excel, vložiť hypertextové odkazy, nakresliť jednoduchý obrázok pomocou príkazov draw - úsečka, úsečka so šípkou, obdĺžnik, elipsa, štetec, vyplnenie oblasti farbou, zvoliť rôzne šírky čiar atď. Vedieť vložiť do textu obrázky s príponami .bmp, .gif, .jpg na ľubovoľné miesto v texte. Napísať matematické vzorce - suma, integrál, logické operátory, mocniny, horné a dolné indexy, odmocniny, riadkové a stĺpcové vektory, matice s určeným počtom riadkov a stĺpcov.