

Managementul informației

Mihail Radu Solcan

3 octombrie 2007

Cuprins

1	Prezentare generală	6
1.1	Managementul informațiilor în contextul managementului cunoașterii	9
1.2	Impactul social al managementului informației	12
1.3	Managementul ideilor tale	13
1.4	Obiectivele unităților de învățare	14
1.5	Prezentarea generală: termeni-cheie	16
1.6	Prezentarea generală: exerciții	16
2	De la textul clasic la cel electronic	19
2.1	Fișele clasice de lectură	20
2.1.1	Însemnările electronice	23
2.2	Managementul și analiza cazurilor	27
2.3	De la textul clasic la cel electronic: termeni-cheie	32
2.4	De la textul clasic la cel electronic: exerciții și analize de caz	32
3	Extragerea informațiilor	34
3.1	Căutarea informațiilor	35
3.2	Limbaajul căutării informațiilor	39
3.3	Caută și înlocuiește	41
3.4	Sortarea informațiilor extrase	44
3.5	Textele ca surse de date	48
3.6	Caută și verifică	51
3.6.1	Dicționar pentru corectorul ortografic	52
3.6.2	Dicționarele bilingve	54
3.7	Extragerea informațiilor: termeni-cheie	56
3.8	Extragerea informațiilor: exerciții și analize de caz	56

CUPRINS

4	XML și generalizarea sistemului de marcare a textului	59
4.1	Generalizarea marcajelor	60
4.2	Exemple de aplicații care utilizează XML	62
4.2.1	Situl Perseus	62
4.2.2	DocBook	64
4.2.3	Formatul OpenDocument	65
4.2.4	RSS	66
4.3	Avantajele și dezavantajele XML	66
4.4	XML și generalizarea sistemului de marcare a textului: termeni-cheie	67
4.5	XML și generalizarea sistemului de marcare a textului: exerciții și analize de caz	68
5	SQL din perspectiva fluxului informațiilor	69
5.1	SQL pe computerul de acasă	70
5.2	Un exemplu de bază de date SQLite	71
5.3	Limbajele de calculator și fluența circulației informațiilor	75
5.4	Căutarea în baza de date	76
5.5	Ordonarea informațiilor	78
5.6	Prezentarea informației din tabelele SQLite	79
5.7	SQL din perspectiva fluxului informațiilor: termeni-cheie	83
5.8	SQL din perspectiva fluxului informațiilor: exerciții și analize de caz	83
6	LaTeX ca mod de organizare a scrierii unui text	85
6.1	Sirenele din computer	86
6.1.1	Evitarea sirenelor	88
6.1.2	Problema efectelor de rețea	90
6.2	A scrie cu gândul la text	92
6.2.1	Clarificarea obiectivelor	92
6.2.2	Stăpânirea structurii	94
6.2.3	Atenția acordată auditoriului	95
6.2.4	Cărămizile textului	96
6.2.5	Trimiterile	97
6.3	Limbajul tehnoredactării	99
6.4	LaTeX ca mod de organizare a scrierii unui text: termeni- cheie	100

CUPRINS

6.5	LaTeX ca mod de organizare a scrierii unui text: exerciții și analize de caz	100
7	Etica și managementul informației	103
7.1	Invizibilitatea tehnologiei informației	104
7.2	Varietatea licențelor	107
7.3	Implicații etice pentru managementul informației . . .	109
7.4	Etica și managementul informației: termeni-cheie . . .	110
7.5	Etica și managementul informației: exerciții și analize de caz	110
	Bibliografie	114

Lista figurilor

2.1	Fișă de lectură din Platon	21
2.2	Cazul anticarului	29
3.1	Căutarea termenului „management“	37
3.2	Rânduri cu șirul de semne „science“	38
3.3	Eliminarea unui spațiu alb	42
3.4	Lărgirea contextului căutării	43
3.5	Propoziții pe rânduri logice	44
3.6	Apelarea <i>aspell</i>	53
3.7	Căutare în dicționarul român-englez al Radei Mihalcea cu ajutorul DING	55
4.1	http://www.perseus.tufts.edu/hopper/	63
4.2	Căutare în textul Odiseei	63
4.3	Document scris în DocBook	64
4.4	Un fișier odt dezarhivat	65
4.5	Fișier cu conținutul textului	66
4.6	RSS în Firefox	67
5.1	Fișierele SQLite nu sunt de tip text	72
5.2	Fișier pentru crearea unei baze de date SQLite	73
5.3	Fișier folosit la introducerea datelor	73
5.4	Tabel cu cărți	74
5.5	Afișarea automată a informațiilor din baza de date	81
5.6	Tabel cu cărți pe Internet	81
6.1	Cuprinsul unui proiect	87
6.2	Fragmente din conținutul unui proiect	88
6.3	Proiect în LaTeX	89
6.4	Logo-ul LaTeX	90

LISTA FIGURILOR

6.5	Enunțarea obiectivelor textului	93
-----	---	----

UNITATEA DE ÎNVĂȚARE 1

Prezentare generală

Managementul informațiilor nu este altceva decât o cupolă sub care sunt reunite preocupări mai vechi sau mai noi. O veche asemenea în-deletnicire este informatica documentară.¹ Conducerile firmelor, fie că i-au zis sau nu management al informațiilor, au trebuie să rezolve problema transmiterii informațiilor potrivite la oamenii cărora le sunt destinate și nu oricum, ci la momentul potrivit și în contextul în care respectivii au nevoie de ele. De asemenea, în condițiile folosirii computerelor și rețelelor de calculatoare, operarea cu informații capătă noi dimensiuni.

Cu toții suntem confrunțați cu ceea ce este, conform unei metafore des uzitate, „un ocean de informații“. Cum să găsim ceva în acest ocean? Are ceea ce găsim vreo *relevanță* pentru problema pe care vrem s-o rezolvăm? Acestea sunt genurile de întrebări care conduc la ideea că trebuie să avem un *management al informațiilor*.

În limba română, cuvântul „management“ n-are rădăcini îndelungate în timp și, lipsind o istorie lungă, lipsesc și analogiile, intuițiile care însoțesc un cuvânt vechi. Fortând un pic limba, trebuie să facem însă o legătură cu o activitate. În sensul acesta, cu toții *gospodărim*, inclusiv informații. O foarte scurtă introducere, lesne disponibilă, în ceea ce am putea denumi *gospodărirea informațiilor* este cea a lui D.M.Griffiths.²

Vom expune pe scurt ideile din textul lui Griffiths și apoi vom analiza

¹Această disciplină s-a conturat cu mult înainte ca „informatica“ să fie un nume pentru *știința calculatoarelor*. Puteți afla mai multe despre informatica documentară citind manualul lui Mihailov, Ceornîi și Ghilearevski[6].Munca din cadrul bibliotecilor este una care ține de informatica documentară, dar nu numai ea. Multe firme au sectoare de informatică documentară. Clasice sunt ziarele. Ele stochează informații despre personalități, evenimente etc. Fără aceste informații ar trebui să se rezume la știri și comentarii, fără a putea oferi un suport documentar mai amplu articolelor lor.

²Managing Information - A Practical Guide[4].

1. Prezentare generală

diferențele de perspectivă în raport cu ceea ce ne propunem să oferim aici. Pentru Griffiths fundamentale sunt trei întrebări: De ce informații avem nevoie? Când avem nevoie de ele? Care este acuratețea informației de care avem nevoie?³ Aceste întrebări au un sens doar dacă le privim prin prisma unei decizii. Pentru a lua o hotărâre trebuie recurs doar la informația *relevantă*. Informația respectivă trebuie să fie disponibilă *la timp*. Gradul de acuratețe trebuie să fie cel de care este nevoie pentru a lua decizia.⁴ Pentru Griffiths *informație* este orice stimul care ajunge la creier. Iar pentru a lua decizii aceste stimulări, pe lângă faptul că trebuie să fie relevante, să vină la timp și să fie doar atât de precise cât este nevoie, trebuie să mai fie complete și simple.⁵ Punctul forte al textului lui Griffiths îl constituie exemplele. Vrem să creștem vânzările de înghețată cu șase la sută? Să ne uităm atunci la deciziile ce trebuie luate pentru a atinge acest obiectiv: decizii cu privire la publicitate, lansarea unui nou produs ș.a.m.d. Pentru a lua o decizie cu privire la publicitate trebuie să avem informații despre legătura dintre campania publicitară și vânzări, despre publicitatea pe care o face concurența produselor sale etc.⁶

Meritul unui text precum cel al lui Griffiths stă în caracterul tranșant al afirmațiilor. În felul acesta poți clarifica momentele unde ar exista motive de diferențiere. Nu este greu să fim de acord cu problema: a gospodărirea informațiilor este un mod de a nu ne scufunda în oceanul tot mai vast al informațiilor. Dar pe cine vizează „noi”? Griffiths dă și exemple din viața cotidiană, dar *noi* este *șefa/șeful* dintr-o firmă organizată ierarhic. Șefa cere informații de la subordonați. Ce-i drept, în viața de zi cu zi, se-ntâmplă să fim și proprii noștri șefi și exemplele date de Griffiths merg. Perspectiva rămâne însă cea *de sus în jos*. Șefa/șeful cer informații. Cum sunt însă produse aceste informații? Cum arată lucrurile *de jos în sus*? De unde să știe cele/cei care strâng informațiile ce anume să adune? Este și din această perspectivă nevoie de un management al informațiilor.

Un alt moment este cel al focalizării pe deciziile luate în vederea atingerii unui obiectiv. De cele mai multe ori trebuie însă să atingem un obiectiv nu ca persoane izolate, ca șefe/șefi care comandă, ci într-un proces de cooperare. Să zicem că obiectivul este acela de a aștepta pe cineva la gară. Persoana care așteaptă și cea care este așteptată trebuie să *comunique* informații pentru a-și *coordona* acțiunile.

Griffiths pune pe bună dreptate accent pe *relevanța* informațiilor,

³ *Managing information*[4, p.2] (trimiterile sunt la paginile versiunii pdf a textului).

⁴ *Ibidem*.

⁵ *Managing information*[4, p.3].

⁶ *Managing information*[4, p.10]

1. Prezentare generală

dar presupune că obiectivul de atins este mereu prezent. Există un exemplu clasic de management al informațiilor în care obiectivele ultime nu sunt decât potențiale. O bibliotecă există pentru a deservi cititoare sau cititori. Desigur că aceștia trebuie să ia decizia dacă să se uite sau nu printr-o carte sau alta. Pentru aceasta se alcătuiesc fișele bibliografice; acestea cuprind informații, exact așa cum zice Griffiths, relevante și totodată simple. Ele sunt stocate însă în vederea unui acțiunii potențiale: sunt acele informații de care *ar avea* nevoie cine ar vrea să decidă dacă se uită sau nu prin carte. S-ar putea însă ca nimeni să nu se uite niciodată printr-o carte sau alta. Iar, mai mult decât atât, principala problemă a unei biblioteci nu este atât furnizarea informațiilor bibliografice, cât *clasificarea* lor. Fără clasificare ne-am rătăci într-o bibliotecă uriașă.⁷

Ideea că informațiile sunt stimulările care ajung la creier nu este doar prea largă. Este greșită. Ea leagă informația de prezența creierului omului. Lucrurile stau exact pe dos. Gestiunea mărfurilor - o activitate cât se poate de prozaică - a fost una dintre ocupațiile care i-au condus pe oameni la descoperirea avantajelor folosirii documentelor scrise. Cea mai veche formă scrisă cunoscută a limbii grecești, linearul B, ne-a parvenit grație în special acestor însemnări ale gestionarilor de depozite.⁸ Informațiile au tocmai această calitate de a „supraviețui“ independent de creierul uman. Ele se găsesc în cărți, în tot felul de evidențe contabile, în alte documente și, mai nou, desigur în computere.

Tocmai impactul computerelor asupra stocării și prelucrării informațiilor este cel care lipsește în textul lui Griffiths. Internetul este pomenit într-un singur alineat.⁹ Griffiths atrage atenția asupra pericolului de a lua o decizie pe baza unor informații învechite de pe Internet. Pericolul este real, dar tocmai existența lui arată tendința tot mai răspândită de a folosi informații de pe Internet.

Biblioteca și Internetul vor fi, prin contrast cu textul lui Griffiths, în centrul atenției aici. Prelucrarea informației cu ajutorul calculatorului și comunicarea ei prin Internet au ajuns un fenomen tot mai răspândit. Sunt puține localitățile ceva mai mari care să nu aibă măcar un Internet

⁷Autorul acestor rânduri tocmai „ieșise în lume“ în 1990 și a descoperit că imensa bibliotecă a Universității Cambridge este cu acces la rafturi (lucru neobișnuit la noi atunci). Nici clasificarea nu te ajută să te descurci. Biblioteca este o adevărată junglă: sunt peste cinci milioane de cărți. La un moment dat doar ghidajul oferit cu generozitate de filosoful australian Graham Priest i-a permis autorului să găsească publicația căutată.

⁸Pentru mai multe informații despre linearul B a se vedea însemnările autorului despre „Descifrarea Linearului B ca exemplu de cercetare științifică“ <http://www.ub-filosofie.ro/~solcan/FSU/c05_06/minos.html>.

⁹*Managing information*[4, p.19].

1. Prezentare generală

café. La nivelul Uniunii Europene răspândirea Internetului este deosebit de mare.

Punctul de vedere adoptat aici este că nu trebuie să *fetișizăm* însă tehnologia informației. N-are noimă să spunem „așa vrea computerul“. Computerul este o unealtă valoroasă pentru că este o unealtă *universală*: prin programare poate fi realizat *orice* mod de prelucrare a informației. Dependența de un anumit mod de lucru, de anumite programe nu este nicicum o fatalitate tehnologică. Multe din deciziile care par de ordin tehnologic sunt decizii ce țin de management, de *managementul informațiilor*.

Fetișizarea tehnologiei informației a dus la un cult straniu al „puterii“ computerului. Oamenii vor computere „puternice“. În focul dezbaterii din Parlament privitoare la laptopul pentru copii, un manager relatează într-un comentariu pe o pagină de Internet cum unul dintre angajații săi a venit să-i ceară să-i dea și lui un computer puternic. Analizând situația, managerul a constatat că angajatul nu reușise de fapt să învețe o combinație de taste. Folosind combinația respectivă de taste computerul ar fi devenit „puternic“.

1.1 Managementul informațiilor în contextul managementului cunoașterii

Ce înțeles are termenul „informație“? Ce *este* informația? Întrebările acestea sunt puse de la primele întâlniri de la seminariile de managementul informațiilor. Este tocmai momentul să demonstrăm cum putem folosi chiar managementul informațiilor pentru a afla ce este *informația*.

Prima reacție ar fi să căutăm o definiție. De pildă, un dicționar de informatică documentară definește informația drept „conținutul oricărei comunicări“¹⁰. Caracterizarea aceasta este foarte generală: conținutul unei comunicări prezintă interes și pentru filosofie, dar și pentru lingvistică, știința computerelor, management etc.

Închipuiți-vă că lucrați în sectorul de documentare al firmei, vă cheamă șeful și vă cere să-i aduceți peste câteva zile elementele necesare pentru a-și clarifica ideea de *informație*. În sine exemplul pare curios. Ce materiale să-i aduceți? Chiar nu știe ce este o informație? Dacă-i veți aduce definiția de mai sus, normal ar fi să se dispenseze după o vreme de colaborarea cu dumneavoastră.

¹⁰ *Dicționar terminologic de informatică documentară* [în rusă, bulgară, maghiară, spaniolă, macedoneană, germană, polonă, română, sârbo-croată, slovacă, slovenă, cehă, engleză și franceză] (Moscova, 1975), definiția 635.

1. Prezentare generală

Să încercăm însă să gândim din perspectiva managementului informațiilor. Șefului i-ar folosi să știe în ce publicații periodice se scrie despre informații. De asemenea, ar fi util să afle cine sunt autoarele/autorii care scriu despre informații, cine-i citează sau ce manuale și enciclopedii se referă la ele/ei; sau ce fișe de post au cei care au studiat manuale despre informații. Acum este însă limpede de exemplu este curios. Ar fi curios să ceară cineva o documentație atât de amplă.

Ce s-ar întâmpla dacă ipoteticul șef ar fi interesat de „managementul informațiilor”? Managementul este o disciplină, ca să zicem așa, „regională”. Chiar și în epoca globalizării. Nu este matematica. Există școli diferite, abordări diferite. Avem desigur de ales între a prezenta o anumită tendință sau a încerca să găsim zona în care ele se suprapun.

Putem începe prin a ne face o idee despre profilul publicațiilor periodice. De pildă, *Information & Management* încearcă să se adreseze deopotrivă celor care implementează sisteme informatice și managerilor.¹¹ *International Journal of Information Management* acoperă un spectru vast de teme, inclusiv crearea de informații și cunoaștere, managementul cunoașterii, abordări filosofice ale cercetării managementului informațiilor.

Așa cum ne-am aștepta managementul informațiilor este o preocupare care precede era computerelor personale și a Internetului. În fond, fiecare universitate, agenție guvernamentală, firmă mai mare are o arhivă. În arhivă este mai multă sau mai puțină ordine. Cine are grijă de ea a fost sau la cursuri de specialitate. A învățat poate acolo mai mult chestiuni teoretice, dar cum-necum arhiva funcționează. Și este una din sursele de informații pentru conducerea agenției sau a firmei, pentru clienți.

În țările de limbă engleză se spune că firmele păstrează *records*, documente (pe hârtie și, mai nou, electronice) ale firmei. Association of Records Managers and Administrators (ARMA) a fost înființată în 1955, înainte deci de era computerelor personale. Azi ea se concentrează desigur asupra documentelor în format electronic.

Dar care sunt legăturile cu managementul cunoașterii? Cum le-am putea descoperi? Arhiviste și arhiviști au fost de multă vreme cei care au gestionat ceea-ce-se-știe în cadrul firmei.¹² La fel, documentariștii din universități sau biblioteci, bibliografele au gestionat în fond ceea-ce-se-știe, cunoașterea.

¹¹ Aflarea adreselor de Internet ale revistelor și altor surse de informații este lăsată ca exercițiu pentru cititoare/cititori. A se vedea secțiunea „Exerciții”.

¹² „Archivists and records managers have long been knowledge managers” - scrie Elizabeth Yakel, în „Knowledge Management: The Archivist's and Records Manager's Perspective”, în *Information Management Journal* (1 iulie 2000).

1. Prezentare generală

Astăzi se pune un accent tot mai mare pe abordările interdisciplinare. O asemenea perspectivă interdisciplinară oferă și managementul cunoașterii. William Saffady arată că „nu obținem cunoaștere doar prin simpla intrare în posesia unui document sau accesând o bază de date”¹³.

În esență, din perspectiva managementul cunoașterea se distinge prin legătura ei constantă cu mintea umană. Prin contrast, și un sistem computerizat poate să intre în posesia unor informații și să le prelucreze.

Dacă investigați literatura de specialitate, veți descoperi că asocierea investigației cunoașterii cu sistemele informatice duce și la crearea unei literaturi impregnate de conceptele și tehnicile științei computerelor. Astfel, dacă vom consulta revista *Knowledge and Information Systems*, vom găsi tocmai articole scrise în acest stil. Revista aceasta și altele cu același profil sunt însă și semnul unei asocieri strânse între managementul cunoașterii și managementul informației.

În contextul unei abordării interdisciplinare, proprii managementului cunoașterii și managementului informației, este nevoie de un reper, de un ax în jurul căruia să se dispună aceste unități de învățare a managementului informațiilor. Opțiunea noastră este pentru axul „cunoaștere și sisteme de prelucrare și transmitere a informației” (sugerat desigur și de denumiri precum *Knowledge and Information Systems*). Atenția noastră se îndreaptă deci asupra acelei părți a cunoașterii care poate fi transformată în informație, prelucrabilă de către mașinile de calcul. Prelucrarea aceasta nu o înțelegem însă prin prisma unui fatalism tehnologic. Tehnologia oferă un evantai de posibilități. Între aceste posibilități este de ales ținând cont tocmai de *factorul uman* - aici și intervine ideea combinației dintre management și informații. Managementul nu ca simplă administrare a unui sistem tehnic, ci ca o modalitate prin care oamenii sunt cei care stabilesc obiectivele, nu sistemul tehnic, sunt cei care determină organizarea și care folosesc sistemele tehnice pentru formarea și dezvoltarea lor ca persoane.¹⁴

Vom mai spune pe parcursul acestor unități de învățare că tehnologia computerelor este și o șansă, și un blestem. Un bun exemplu în acest sens sunt popularele complexe de tip *office*. Un program precum Word a democratizat accesul la tehnologia informației. Un sistem clar de meniuri și posibilitatea de a seta o serie de opțiuni relativ simple au făcut posibilă abandonarea mașinii de scris și înlocuirea ei cu un dispozitiv radical diferit din punct de vedere calitativ. În același timp, sistemul are limitele sale: n-are cum să suplinească învățarea unui lim-

¹³William Saffady, „Knowledge Management: A Manager’s Briefing”, în *Information Management Journal* (1 iulie 2000).

¹⁴Compară și cu Peter Drucker[1, p.400].

1. Prezentare generală

baj de programare (care deschide cu adevărat calea către posibilitățile nelimitate ale computerului), conduce către o tiranie a vizualului (în dauna organizării logice a textului) și stimulează fetișizarea tehnologiei (ajungi să crezi, de pildă, că tehnoredactarea simultană nu este doar o posibilitate oferită de tehnologie, ci o fatalitate).

1.2 Impactul social al managementului informației

Să facem un experiment imaginar. Să zicem că lucrăm la o mare bibliotecă și menirea noastră este să clasificăm cărți. Cu alte cuvinte, trebuie să producem informațiile bibliografice despre carte și să spunem care este domeniul creației umane în care se încadrează cartea. N-am auzit niciodată de un autor pe nume Thomas Kuhn. Cartea pe care o avem de prelucrat se numește *Tensiunea esențială* (București: Editura Științifică și Enciclopedică, 1982). Dacă se vorbește în titlu despre „tensiune“ este aceasta cumva o carte de electronică, de electrotehnică? Trebuie oare să citim toată cartea sau studiul introductiv pentru a o putea clasifica? Pentru a rezolva problema o metodă extrem de eficientă este examinarea citatelor. Pe cine citează Kuhn? Examinarea citărilor va elimina imediat bănuiala că ar putea fi vorba despre o carte de electrotehnică. Este vorba de o carte de filosofie, cu deschideri și către alte domenii (în special către istoria științei).

Metoda examinării citărilor este una puternică și o putem aplica și unui articol despre rețelele de citări.¹⁵ În lista de referințe bibliografice, la numerele 14 și 15, găsim numele lui Eugene Garfield. Dacă ne uităm atent la anii când au apărut articolele sale, 1955 și 1972, vedem că studiile lui Garfield sunt considerabil mai vechi decât celelalte articole. Garfield se află, de fapt, la originea ideii de a folosi rețele de citări pentru a investiga literatura științifică.

În 1960 Eugene Garfield a fondat Institute for Scientific Information (ISI)¹⁶. Aceasta este organizația care publică indici ai citărilor. Cu ajutorul acestor indici se poate aprecia impactul unei lucrări și al scrierilor unei autoare/autor.

Un index al citărilor poate sluji mai multor scopuri. Este, în primul rând, un instrument de investigare științifică a comunităților academice. În aceste comunități, grație citării într-o lucrare a altor lucrări,

¹⁵Wangzhong Lu, J. Janssen, E. Milios, N. Japkowicz, Yongzheng Zhang, „Node similarity in the citation graph“, *Knowledge and Information Systems*, vol.11, nr.1, 2006, pp. 105–129.

¹⁶Astăzi Thomson Scientific <<http://scientific.thomson.com/>>.

1. Prezentare generală

se pot stabili legături, se pot trasa filiații ale ideilor, se poate vedea opera cui a avut un impact mai mare.

Metoda urmăririi citărilor este extrem de utilă și-n managementul informației. Foloasele pe care le poate aduce se văd chiar la nivelul alcătuirii unei bibliografii de mici dimensiuni (pentru o lucrare la un curs/seminar, pentru o disertație). N-avem decât să urmărim firele citărilor pentru a realiza bibliografia de care avem nevoie. Trebuie desigur să fim atente/atenți la rolul unei lucrări sau a alteia. Dacă - precum în exemplul de mai sus - o lucrare are un rol fundamental în domeniu, nu trebuie s-o oțimem. Concret, pornind de la exemplul dat, nu trebuie să-l oțimem pe Garfield. Mai mult decât atât, trebuie să identificăm și alte lucrări care-l citează. Invers, lucrări mai puțin citate, periferice sau aflate oricum în afara tematicii investigației noastre, trebuie omise. Deci, oarecum paradoxal, în contextul exemplului nostru, am putea omite articolul de la care am pornit (de pildă, dacă nu ne ocupăm cu algoritmi care stabilesc similarități între grafuri ale citărilor).

Ceea ce a început ca o cercetare/management al informațiilor a avut și un considerabil impact asupra comunităților științifice. A devenit tot mai important să apară într-o publicație indexată de către ISI. De asemenea, munca de cercetare ca atare a început să fie apreciată cu ajutorul indexului citărilor.

Impactul indicilor ISI este considerabil. Au apărut și multe discuții cu privire la valoarea relativă a citărilor. Mai nou ISI a pătruns și-n viața comunităților academice de la noi din țară. Primul efect (deja simțit) va fi acela al încercării de a obține indexarea din partea ISI a cât mai multor reviste interne. Până acum, multe dintre le pur și simplu nu erau indexate de ISI.

1.3 Managementul ideilor tale

Peter Drucker (1909-2005) este poate cel mai cunoscut cercetător al fenomenului managementului din secolul trecut. Este socotit de mulți părintele managementului ca disciplină academică.

Într-un articol din anul 2000, Peter Drucker afirma tranșant că până prin 1900 nimeni n-avea opțiuni; n-aveai de ales, dacă erai fiică sau fiu de țărani, tot țăran rămâneai, dacă tata era meșteșugar, în marea majoritate a cazurilor, tot meșteșugar ajungeai. Secolul al XX-lea a adus posibilitatea de a alege între cariere variate, susține Drucker. Mai mult decât atât, a devenit posibil să nu practici aceeași meserie toată viața.¹⁷

¹⁷Peter Drucker, „Managing Knowledge Means Managing Oneself“[2].

1. Prezentare generală

În fapt, susține Drucker, cunoașterea este cea care creează șanse. Tot cunoașterea explică de ce sunt acum femei pe posturi ocupate odinioară doar de bărbați.

Mutatis mutandis, ideile lui Drucker despre lucrătorul care folosește cunoașterea de care dispune ne permit să-nțelegem avantajele pe care le putem obține dacă organizăm mai bine modul în care lucrăm cu ideile noastre. Marii creatori invocați de Drucker au fost și oameni care au știut cum să practice managementul gândurilor lor.

În acest context, sistemele informatice moderne ne oferă tuturor o șansă unică. Computerul este o unealtă extrem de flexibilă și poate fi folosit pentru un management mult mai eficient al ideilor noastre - pe calea managementului informațiilor.

Condiția pentru un management mai eficient al ideilor noastre este însă să nu ne lăsăm dominați de tehnologie. Mai corect spus, să nu lăsăm pe alții, cei care au creat programele de calculator pe care le folosim, să aleagă în locul nostru.

1.4 Obiectivele unităților de învățare

Un lucru extrem de recomandabil ar fi ca, înainte de a citi această secțiune, să rezolvați exercițiile din această unitate de învățare. Din perspectiva managementului informației, ele sunt foarte simple. Pornesc de la ceea ce noi toți facem cel mai adesea acum: căutăm informații pe Internet.

Întrebarea ar fi însă la ce nivel ar fi rezonabil să plasăm obiectivele acestor unități de învățare. La nivelul relativ simplu amintit mai sus sau la nivelul articolelor dintr-o revistă cu un puternic accent pe modul de funcționare al sistemelor informatice. Trebuie găsit evident un echilibru.

Trebuie spus un cuvânt despre programele de calculator menționate în continuare. Realist vorbind, este imposibil de conceput managementul informației, în secolul al XXI-lea, fără computere. Chiar dacă s-ar putea ca, la lectură, cititoarea sau cititorul să socotească unele pasaje drept prea tehnice, obiectivele sunt alese în așa fel încât să vizeze dobândirea de cunoștințe fundamentale. În special când este vorba despre programele de calculator oamenii sunt obișnuiți să li se furnizeze cunoștințe despre *utilizarea* lor. Aici nu este vizată utilizarea, ci înțelegerea noțiunilor fundamentale legate de *pașii* care pot fi făcuți cu ajutorul calculatorului.

Ca-n orice scriere, contează și experiența autorului. Experiența practică a autorului este în domeniul informaticii documentare. De aici

1. Prezentare generală

tendința desigur de a da mai ales exemple cu texte, cărți și biblioteci. Lipsesc evidențele contabile etc., dar cititoarele și cititorii sunt invitați să construiască analize de caz pe baza propriei experiențe.

Auditoriul inițial al cursului provenea în mare parte din zona celor care construiesc și întrețin sisteme informatice. De aceea, prima versiune cuprinde sensibil mai multe exemple legate de sistemele informatice. Între timp auditoriul s-a schimbat și această versiune a cursului încearcă să țină cont de această mutație. Exemplele legate de sistemele informatice sunt alese din zone de care oricine s-ar lovi dacă deschide un computer.

Toate programele la care ne referim în text sau care sunt indicate în exerciții pot fi descărcate de pe Internet. Banii nu sunt deloc o piedică în procurarea lor. De asemenea, fiind programe cu sursă deschisă infectarea lor cu programe distructive („virusi“) este practic exclusă. Nu le descărcați însă decât de pe sursele cu software cu sursă deschisă. Chiar dacă dumneavoastră v-ar veni greu să cercetați sursele, sunt suficienți alți ochi care s-au uitat peste sursele respective.

Obiectivele cursului sunt:

- Să prezinte managementul informațiilor în strânsă legătură cu uneltele de care dispunem grație computerelor;
- Să introducă variante electronice ale textului fișelor de lectură și al analizelor de caz;
- Dobândirea unei cunoașteri a limbajului căutării, substituției și sortării de informații dintr-un text;
- Însușirea fundamentelor limbajelor de marcare a textelor și a reprezentării structurate a informațiilor;
- Înțelegerea modului în care un limbaj pentru baze de date asigură continuitatea fluxului informațiilor (de la sursele informațiilor până la destinație, inclusiv pe Internet);
- Prezentarea informațiilor într-un mod rațional în procesul de redactare și tehnoredactare a textelor (rapoarte, proiecte, lucrări, disertații etc.);
- Discutarea aspectelor ce țin de etică în managementul informațiilor.

Sunt șapte unități de învățare a managementului informațiilor. Metaforic vorbind, unitățile 3-6 alcătuiesc nucleul „tare“, iar unitățile 1-2, 7 formează învelișul „moale“. Pentru firul-conducător al unităților de

1. Prezentare generală

Învățare a managementului informațiilor citiți secțiunea „Limbașele de calculator și fluența circulației informațiilor“ din unitatea de învățare „SQL din perspectiva fluxului informațiilor“. Cuvintele-cheie din denumirea secțiunii sunt „limbaș“ și „fluență“. Limbașul asigură flexibilitate în atingerea obiectivelor și eficiență. Ar fi însă inutilă eficiența „pe bucăți“. Întregul curs al informației, de la sursele ei la cine beneficiază de ea, trebuie să fie fluent.

1.5 Prezentarea generală: termeni-cheie

- informație
- relevanța informației
- sistem informațional
- managementul cunoașterii
- managementul informației

1.6 Prezentarea generală: exerciții

Exercițiul 1: În multe din exercițiile de la unitățile de învățare a managementului informațiilor va fi folosită expresia „folosiți ca punct de plecare articolul... din Wikipedia“. Trimiterea este la un articol în limba engleză. Ce faceți dacă nu știți această limbă? Cu titlul de exercițiu descoperiți articole similare unor articole în limba engleză. Folosiți butonul „Random article“ pentru a găsi articole în mod aleator. Adresa Wikipediei în engleză este <<http://en.wikipedia.org/>>.

Exercițiul 2: Wikipedia este scrisă de voluntari. Nu are redactori profesioniști. Oricine poate schimba conținutul articolelor. Ideea este că, asemenea prețului pe o piață liberă, conținutul articolelor ajunge la o formă optimă în urma unor repetate „tranzacții“ (cine schimbă articolul oferă o nouă formă și, implicit sau explicit, negociază cu alte persoane care vor să contribuie la articol). S-ar putea să nu găsiți articolul ajuns la forma sa optimă (mai ales atunci când numărul tranzacțiilor cu articolul respectiv a fost mic). Verificați întotdeauna orice informație folosind cel puțin încă o sursă, independentă de cea din Wikipedia. Faceți acest lucru cu câteva articole alese la întâmplare.

Exercițiul 3: Wikipedia este un fenomen al secolului al XXI-lea. Pentru mulți este un gen de enciclopedie în care n-au încredere din principiu. Din motive uneori de ordin diplomatic, nu este bine să citați

1. Prezentare generală

Wikipedia. Urmați însă îndemnul „folosiți ca punct de plecare Wikipedia“. Faceți acest lucru cu câteva articole alese la întâmplare. Vedeți la ce surse de informații, altele decât Wikipedia, puteți ajunge.

Exercițiul 4: Accesați situl Editurii Elsevier, una dintre marile edituri academice, la adresa <<http://www.elsevier.com/>>. Introduceți în caseta de căutare termenul "information management" (cu tot cu ghilimele). Cercetați profilul revistelor (intrările marcate cu „journal“). Apăsați pe meniul din stânga rubrica „most downloaded articles“. Citiți unele rezumate.

Exercițiul 5: Porniți de la articolul despre ARMA din Wikipedia <http://en.wikipedia.org/wiki/Association_of_Records_Managers_and_Administrators>. Aflați mai multe despre asociație. Cercetați profilul revistei publicate de asociație.

Exercițiul 6: Accesați situl <<http://www.thefreelibrary.com/>>. Mergeți la categoria *business*. Căutați *Information Management Journal*. Mergeți la numărul din 1 iulie 2000. Citiți despre *knowledge management* (managementul cunoașterii).

Exercițiul 7: Accesați pe Internet situl <<http://www.igi-pub.com/journals/>>. Căutați *International Journal of Knowledge Management*. Investigați profilul publicației.

Exercițiul 8: De multe ori este imposibil pentru o bibliotecă să se aboneze la toate revistele. Cu atât mai puțin este posibil acest lucru pentru o persoană privată. Cercetați bazele de date la care este abonată biblioteca la care aveți acces. De pildă, dacă studiați la Universitatea din București și frecvențați Biblioteca Centrală Universitară puteți folosi baza de date a Editurii Springer. Adresa pe Internet este <<http://www.springerlink.com/>>, dar trebuie s-o accesați de la computerele bibliotecii. Căutați reviste al căror titlu începe cu litera „k“. Găsiți *Knowledge and Information Systems*. Studiați profilul revistei. Vedeți în ce condiții puteți citi articolele.

Exercițiul 9: Accesați Google Book Search <<http://books.google.com/>>. Introduceți în caseta de căutare "Thomas Kuhn" (puneți și ghilimelele!). Din lista de cărți scrise de Thomas Kuhn sau în care este vorba despre Thomas Kuhn alegeți Thomas Nickles (editor), *Thomas Kuhn* (Cambridge University Press, 2003). Apăsați „+“ de la meniul „popular passages“. Mergeți, de exemplu, la pagina 41. Examinați lista bibliografică. Faceți-vă o idee mai generală despre lucrările în care este citat Thomas Kuhn. din prezentarea făcută lui Kuhn la volumul editat de Nickles reiese că acesta a fost un filosof. Oferă rețeaua de citări o întregire a acestei caracterizări?

Exercițiul 10: Porniți de la articolul „Citation index“ din Wikipedia <http://en.wikipedia.org/wiki/Citation_index>. Identifi-

1. Prezentare generală

fiți surse de informații despre metoda indexării citatelor. Identificați surse de unde puteți afla cine citează lucrări științifice.

Exercițiul 11: Citiți din Wikipedia articolul „Reputation management“ <http://en.wikipedia.org/wiki/Reputation_management>. Încercați să faceți o comparație între comunitățile descrise în articol și o comunitate științifică. În ce sens un număr mai mare de citări înseamnă un rang mai înalt în ochii comunității.

Exercițiul 12: Citiți din Wikipedia articolul despre Peter Drucker <http://en.wikipedia.org/wiki/Peter_Drucker>. La data de 1 octombrie 2007 acest articol avea o atenționare privitoare la lipsa unor surse credibile ale informațiilor din articol. Verificați informațiile din articol. Încercați să vă faceți o idee apoi despre rolul lui Peter Drucker în evoluția științei managementului.

UNITATEA DE ÎNVĂȚARE 2

De la textul clasic la cel electronic

În această unitate de învățare vom studia două tehnici de muncă intelectuală, atât în varianta lor clasică („cu creionul și hârtia“), cât și în varianta electronică (folosind computerul). Prima dintre cele două tehnici, cea a însemnărilor de lectură, este de interes general, fiind utilizată de către oricine trebuie să elaboreze o lucrare „cu note de subsol și bibliografie“ (o scriere de tip academic). A doua tehnică este specifică doar anumitor tipuri de activitate, aici incluzându-se și managementul în diversele sale forme: studiul sau analiza cazurilor.

Obiectivele acestei unități de învățare sunt:

- Însușirea elementelor unei fișe de lectură simplă, în versiunea sa clasică.
- Înțelegerea deosebirii dintre aspectul vizual al unui text electronic și structura sa logică.
- Înțelegerea contrastului între un editor de texte (orientat către aspectele logice ale textului) și un program destinat tehnoredactării de texte (orientat către aspectul vizual al textului).
- Deprinderea unui mod de a realiza însemnări de lectură în format electronic după o structură standard.
- Deprinderea unui mod de a realiza o analiză de caz respectând un format standard (atât din perspectiva conținutului, cât și a formei textului electronic).

2.1 Fișele clasice de lectură

Modulele acestui manual se numesc „unități de învățare“. Puneți-vă în cea mai simplă, dar și cea mai importantă „situație de învățare“: citiți o carte sau un alt document scris. Nu puteți să vă bazați doar pe memoria umană; limitele ei sunt considerabile. Mai ales dacă vreți să folosiți într-un proiect propriu ceea ce-ați citit va fi extrem de greu să regăsiți informația dacă nu recurgeți la o formă de management al informației citite.

Semnul de carte este mijlocul cel mai simplu de a plasa un reper în text. Semnul de carte nu este însă de mare folos. Este bun dacă vrem să știm până unde am ajuns cu lectura. Dacă pun însă prea multe semne de carte, peste o vreme nu le voi ști rostul.

O idee este să folosim drept semne de carte fâșii de hârtie pe care ne facem niște însemnări. Este un pas înainte. Acum știm ce informație am socotit că este importantă în textul respectiv. Dezavantajul acestui sistem este caracterul său static. Chiar operații simple de prelucrare a informației devin dificile. Încurcătura cea mare provine din faptul că nu pot reordona informațiile găsite fără să pierd legătura cu textul.

Soluția la problema de mai sus este folosirea fișelor de lectură. O fișă de lectură este un exemplu excelent de cunoaștere căreia i-am dat o structură: am aflat ceva și am structurat ceea ce-am aflat, transformând cunoașterea noastră într-o informație care este stocată, dar poate fi și regăsită, sortată ș.a.m.d. Pe scurt, fișa de lectură este cel mai simplu, dar și cel mai important, exemplu de trecere de la cunoaștere difuză la o informație.

Fișele clasice de lectură au formatul unei cărți poștale. Este foarte important să respectăm acest standard atunci când alcătuim fișe de lectură. Dimensiunile diverselor tipuri de fișe sunt rodul unei îndelungate experiențe. N-are rost să refacem această experiență, descoperind pe propria piele defectele altor soluții.¹

Să examinăm un exemplu, normal în cadrul filosofiei, exotic în afara ei, dar care ilustrează înstructurarea unei fișe de lectură simple. Să zicem că vreau să aflu (de la sursă) ce zice Platon despre cunoaștere. Mă apuc să citesc dialogul *Theaitetos* și dau peste un pasaj interesant. Pun pe hârtie (v. figura cu fișa de lectură din Platon) ceea ce-am aflat.

Survin o serie de probleme. În primul rând se vede că am divizat gra-

¹Pentru detalii cu privire la fișele de lectură și alte teme legate de explorarea unei documentații și construirea unui text v. Mihail Radu Solcan, *Eseul filosofic*[11]. Cartea are și un sit pe Internet la adresa <http://www.ub-filosofie.ro/~solcan/eft/index.html>.

2. De la textul clasic la cel electronic

Semnătura	Theaitetos: 201c-d
cunoaștere	Platon cercetează formula conform căreia cunoașterea este crezare adevărată și întemeiată
λόγος	ἔφη δὲ τὴν μὲν μετὰ λόγου ὀληθῆ δόξαν ἐπιστήμην εἶναι, τὴν δὲ ἄλογον ἐκτὸς ἐπιστήμης
	Vezi și Schnädelbach

Figura 2.1: Fișă de lectură din Platon

fic fișa în câmpuri. În colțul din stânga sus punem semnătura noastră.² Celelalte câmpuri trebuie explicate ceva mai amănunțit.

Câmpul din dreapta sus conține date cu privire la sursa informațiilor. Cum n-are sens să scriem de fiecare dată toate datele bibliografice, folosim o prescurtare. Este însă important să avem datele bibliografice referitoare la cartea sau orice alt document care constituie sursa noastră. Dacă avem în biblioteca proprie cartea, putem plasa fișele în carte. Altfel trebuie neapărat să avem datele bibliografice. O metodă simplă este să scriem aceste date pe un plic și să punem în plicul respectiv fișele de lectură.

Tot în câmpul din dreapta sus există și o informație cu privire la locul în care se găsește ideea reținută de noi. În exemplul folosit aici, n-am scris doar „Theaitetos“. Ar fi prea vag. Dialogul lui Platon este mult prea amplu ca să găsim cu ușurință punctul în care se află ideea pe care vrem s-o reținem. Am pus de aceea și pagina standard din Platon. De ce pagina standard? Dacă am pune, să zicem, „p.260“ (a traducerii românești din 1989) n-am putea să folosim fișa și pentru a regăsi informația în alte traduceri sau în textul grecesc al dialogului platonice.

Ce ne facem dacă nu există pagini standard pentru textul respectiv? Soluția este să indicăm segmentul din text: capitolul, secțiunea, alineatul. De multe ori, va trebui să calculăm singuri asemenea numere, dar

²Dacă prezint fișa la un examen este important să fie personalizată. Dacă lucrăm într-un colectiv, la un proiect, este bine să se știe de unde provine fișa respectivă.

2. De la textul clasic la cel electronic

ele sunt cât se poate de utile. Referindu-ne la segmentarea textului, nu la pagini, putem să căutăm informația în orice ediție.

Câmpul principal al fișei de lectură este cel din dreapta jos. Regula de bază este să prezentăm aici o singură idee-cheie. Dacă punem mai multe idei, fișa devine neoperațională. Dacă vrem să regrupăm fișele, una sau mai multe idei nu mai sunt la locul lor.

A pune o singură idee-cheie pe o fișă nu înseamnă să scrii doar două-trei cuvinte. Ideea trebuie prezentată cât mai bine, într-o propoziție completă. De asemenea, are sens să adăugăm exemple sau extrase din text. Totul trebuie însă să graviteze în jurul unei singure idei, nu a mai multora.

În exemplul de mai sus am reținut un citat din text. În mod intenționat el este dat în original. În acest fel este cât se poate de limpede că este vorba despre textul altcuiva. Multă lume plătește un preț greu pentru neglijențele în alcătuirea fișelor de lectură. Dacă, în momentul în care citești, folosești pe fișă o formulare din text, fără a evidenția în mod absolut clar că ea nu-ți aparține, va exista riscul s-o reproduci în lucrarea ta, fără a mai preciza că este vorba despre un citat - ceea ce desigur este o încălcare a eticii academice. Un text în limba originalului este un semn absolut limpede că ideea nu-ți aparține.

Dacă reținem un citat în limba română, trebuie să marcăm cu grijă acest lucru cu ajutorul ghilimelelor. De pildă aș fi putut să-mi notez că, în traducerea românească din 1989, se vorbește despre „opinia adevărată, însoțită de înțelese“(p.260).

Dacă textul n-are pagini standard sau vrem să ne amintim exact la ce pagină dintr-o anumită ediție se găsește citatul, atunci trebuie să notăm în paranteză numărul paginii. Nu este însă, după cum spuneam, deloc recomandabil să inserăm aceste numere de pagină non-standard în câmpul din dreapta sus. Am îngusta în acest fel prea mult aria operațiilor pe care le putem face cu fișa respectivă.

Este, de asemenea, util să includem trimiteri către lucrări în care ideea de pe fișă este discutată. De pildă, concepția despre cunoaștere a lui Platon este discutată de către Herbert Schnädelbach în introducerea sa în teoria cunoașterii[10].

Multă lume simte nevoia să numeroteze fișele. Este o eroare. Numerele își pierd sensul când reordonăm fișele. Trebuie folosite trimiteri de forma „abreviere: locul din structura sursei“. Aceste trimiteri nu-și pierd noima nici dacă sortăm fișele.

Cum să operăm însă sortările fișelor? Folosim cuvintele-cheie, scrise în câmpul din stânga jos. Cuvintele-cheie scot în evidență ceea ce este relevant în informația de pe fișă. În exemplul de mai sus, am reținut

introducerea, de către Platon, în discuția despre cunoaștere, a întemeierii.

2.1.1 Însemnările electronice

În mod deliberat această secțiune nu se numește „Fișe de lectură electronice“ sau ceva de acest gen. Computerul este o binefacere, dar și un blestem. Lucrul acesta se vede cât se poate de bine în managementul informației.

Când a apărut tiparul, meșterii s-au străduit să producă un text cât mai apropiat de acela al manuscriselor. Estetic, pagina tipărită a avut în mod cert de câștigat. Treptat s-a mai renunțat la ligaturi și letrine complicate, dar efortul de a imita manuscrisul se vede până astăzi.

În cazul fișelor de lectură, tentația este aceeași. Apare însă o mare problemă de management. Tehnic - mai ales astăzi - este perfect posibil să imiți o fișă scrisă de mână. Poți să trasezi toate liniile, să atașezi tot felul de însemnări în baloane, ba chiar să faci în așa fel încât fișele să existe virtual pe ecran, să fie deplasate și aranjate cu *mouse*-ul. Efectul estetic este indubitabil, dar fetișizarea tehnicii nu este deloc recomandabilă.

Un principiu de bază al managementului informației este să nu rupi fluxul informațiilor care circulă în organizație. Din perspectivă managerială, nu tehnologia informației este importantă, ci managementul informației, iar aceasta înseamnă oameni la care trebuie să ajungă informația.

Cel mai simplu indiciu al ruperii continuității fluxului informațiilor este nevoia de a tot retasta informațiile. Un alt indiciu sunt operațiunile de „tăiere și lipire“. Informația circulă de la o persoană la alta, dar munca este una de Sisif. Fiecare tot copie, lipește sau - mai rău - tot reintroduce informațiile în calculator.

Un alt motiv al ruperii fluxului informațiilor este reprezentat de incompatibilitățile dintre programele de calculator. De exemplu, ce se vede pe o platformă Windows, nu se vede pe Linux și invers. Formatul diverselor editoare de texte nu este același și așa mai departe.

De unde să luăm lucrurile pentru a înțelege în mod rațional natura acestor probleme? De la text. În ciuda aparențelor, computerele moderne nu sunt atât mașini de făcut calcule, cât mașini de prelucrat informația conținută în texte. În spatele oricărui calcul complicat stă tot un text: textul programului în care sunt specificați pașii care trebuie făcuți și deciziile care trebuie luate pentru a ajunge la rezultat. Cele mai multe date manevrate cu ajutorul calculatorului nu sunt numerice: sunt texte prelucrate foarte sofisticat cu ajutorul calculatorului. Deci

2. De la textul clasic la cel electronic

chiar dacă ne uităm cu ochiul managerului, nu al tehnologului, trebuie să sesizăm importanța textelor.

Nu vă lăsați însă înșelate sau înșelați de ceea ce vedeți pe ecranul calculatorului. În cazul textelor din calculator există o deosebire subtilă între vizual și logic. Pe ecran vedem o serie de semne. Mai multe desigur - și relativ mai greu de produs - dacă este vorba despre scrierea japoneza sau cea chineză. Mai puține, în cazul scrierilor din Europa. În sine, numărul de semne folosite contează mai puțin. Contează mai puțin și analogiile cu vechile tehnici. Zețarii, lucrători din vechile tipografii, lucrau cu niște litere de metal pe care le țineau în sertare compartimentate.³ În computer, „sertarele compartimentate“ au în ele desene.⁴ Fiecare literă, în toate variantele sale, nu-i nimic altceva decât un mic desen. Spre deosebire de meșterul tipograf noi nu putem apuca fizic aceste desene. Le manevrăm cu ajutorul unor coduri. Natura acestor coduri este însă o chestiune pur tehnologică. Din perspectiva care ne interesează aici, putem începe de la distincția dintre vizual și logic.

Vechii meșteri foloseau literele lor de lemn sau de metal ca pe un fel de ștampile cu care realizau desepe pe hârtie. Cu mașina tradițională de scris, dactilografele puneau pe hârtie micile desene tot cu ajutorul unui soi de „ștampile“ metalice. Astăzi, vedem literalmente desenul literelor pe ecran și-l putem apoi realiza pe hârtie - ca și cum uriașa clădire a tipografiei de altă-dată s-ar fi mutat în colțul casei.

Rândurile pe care le vedem pe ecranul computerului creează însă o mare confuzie. Pe de o parte, sunt rândurile vizuale. Când se termină spațiul, se trece „pe rândul“ următor. Rândurile prea lungi sunt greu de urmărit, așa că operația aceasta este de mare ajutor. Este însă utilă pentru *ochiul* omenesc și atât.

În spatele fiecărui semn din textul de pe ecran se află însă un cod; inclusiv în spatele spațiilor albe. Firește, sfârșiturile de rând sunt și ele codate cumva. N-ar avea sens însă să punem de prima dată un cod pentru a marca sfârșitul unui rând vizual. N-avem cum ști dinainte cât este de lat ecranul sau fereastra de pe ecran în care este afișat textul sau pagina. Codurile se pun atunci când se termină un rând

³Denumirile englezești de *lower case* și *upper case* vin de la aceste sertare sau cutii. Literele mici, folosite mai des, erau ținute la îndemână, în *case*-ul de dedesubt (*lower*). Cele de sus, la care ajungeai mai greu, erau destinate majusculilor - folosite mai rar.

⁴Examinați *font*-urile din calculator cu un program mai sofisticat și veți vedea aceste colecții de desene. Puteti folosi, de pildă, FontForge, un program *open source*, disponibil pe Internet la adresa <http://fontforge.sourceforge.net/>. Există, de asemenea, o sumedenie de programe comerciale care pot juca același rol. Programele obișnuite de vizualizare a *font*-urilor nu sunt însă utile. Ele vă dau o idee doar despre forma literelor pe ecran, nu despre „cutiile“ cu desene ca atare.

2. De la textul clasic la cel electronic

logic. Practic, asta înseamnă atunci când apăsăm tasta cel mai adesea denumită „*Enter*“.

Calculatorul poate fi o binefacere, dar și un blestem. Pentru cei mai mulți oameni, când apasă *Enter*, declanșează - fără să știe - un blestem. Au marcat un sfârșit de rând *logic*; au dorit probabil să construiască altfel un rând vizual.

Din câte știu, pe platforma Windows nu există preinstalat un editor sofisticat de texte pure. Editorul de texte este *Notepad*. Pe alte platforme, cum sunt cele Unix, sunt instalate de la bun început editoare de texte puternice precum Vim sau Emacs.⁵ Textele de computer pure se scriu cu editoare precum Notepad, Vim sau Emacs. Nu folosiți sub nici o formă un editor WYSIWG⁶ pentru a produce text curat, fără coduri pentru aspectul *vizual* al textului.

Deși s-ar părea că luăm firul de la capăt, nu vă faceți prea multe iluzii. Tehnologia codării sfârșitului de rând *logic* sub Windows, Mac și Unix este diferită. Detaliile acestei diferențe nu sunt semnificative aici, pentru că nu ne interesează tehnologia ca atare. Din punct de vedere managerial, complicațiile pot să apară dacă nu știm că până și sfârșitul rândurilor logice este codat diferit.⁷ Obstacolul acesta tehnologic trebuie depășit pentru a nu rupe curgerea fluxului informației în sens managerial.

După această punere în gardă, putem să ne apucăm să construim o versiune electronică a fișelor de lectură. Folosim pentru aceasta un editor de texte, gen Notepad. Practic recomandabil este să recurgem însă la Vim sau ceva asemănător.⁸ Principiul fundamental de la care plecăm este acela că orice informații pe care vrem să le folosim cu adevărat trebuie să aibă o structură.

Putem ignora, pe moment cel puțin, câmpul cu semnătura pusă pe fișele clasice („de hârtie“). Reținem deci trei câmpuri. Două dintre aceste câmpuri este firesc să le gândim ca rânduri logice: câmpul pe care

⁵Puteți însă instala ușor Vim sau Emacs și sub Windows. Editorul Vim, la care mă voi referi adesea, se găsește la adresa <<http://www.vim.org/>>.

⁶Prescurtarea vine de la cuvintele englezești care înseamnă „ceea ce vezi (pe ecran) obții (și pe hârtie)“. Acestea sunt editoare folosite - cu mai mult sau mai puțin succes - pentru a *tehnoredacta* texte, cu alte cuvinte, pentru a le aranja în pagină. Aceasta este cu totul altceva decât producerea unui fișier cu text pur, numit în engleză și *flat file*.

⁷Cu Vim, de pildă, aveți însă ușor posibilitatea schimbării formatului fișierului. Prin meniul *Edit* aveți acces la setările care privesc fișierul, inclusiv formatul acestuia. Atenție doar la faptul că meniul combină ceea ce ține de vizual cu ceea ce ține de logica fișierului text.

⁸Atenție la Vim! Este un editor remarcabil, dar folosirea sa e departe de a fi simplă. Pentru un minimum de informații despre folosirea Vim v. și anexele la *Eseul filosofic*[11].

2. De la textul clasic la cel electronic

se află sursa și cel pe care sunt cuvintele-cheie. Ce facem însă cu un câmp mai complex, precum cel principal? Sau cum deosebim între rolul acestor rânduri de text electronic? Vizual, ne slujeam de disponerea în spațiu a câmpurilor.

Ar fi o eroare să ne apucăm să tragem linii și să imităm fișele scrise de mână. Aceasta este o mare greșeală din perspectiva cursivității fluxului informațiilor. Este nu atât o decizie tehnologic proastă, cât una nefericită din perspectivă managerială. Tehnologia este chiar la îndemână - din păcate, într-un fel - pe computerele actuale. Este ca o vrăjitoare rea care a luat chipul unei zâne bune. Nu-i mare lucru să desenezi fișele într-un fel sau altul. Dar este ca și cum te-ai uita o dată la ele și le-ai băga într-un sertar de unde nu le mai scoate nimeni vreodată decât cu prețul reintroducerii informațiilor în circuit.

Pentru a deosebi între rolul rândurilor logice există o metodă simplă: să le dăm un nume. Estetic nu arată deloc bine, dar ideea este bună pentru managementul informațiilor. Numele rândurilor îl punem la început de rând și-l despărțim de rest printr-un semn căruia îi rezervăm doar rolul de separator. Câmpul principal îl putem sparge în mai multe rânduri logice, fiecare denumit adecvat.

Fișa noastră din dialogul lui Platon *Theaitetos* ar arăta atunci în felul următor:

Sursa :: *Theaitetos* 201c-d.

Cuvinte-cheie :: cunoaștere, logos.

Idee-cheie :: Platon cercetează formula conform căreia cunoașterea este crezare adevărată și întemeiată.

Citat :: ...

Trimitere :: Vezi și Schnädelbach.

Drept separator este folosită perechea de „două puncte“ pentru a evita blocarea folosirii celor două puncte. S-ar putea să avem nevoie de două puncte altundeva (de pildă, în textul ideii-cheie). Dacă am pune două puncte, în același rând logic, în poziții diferite, n-am ști însă care este numele câmpului și care este conținutul.

Primele trei rânduri logice sunt obligatorii. Celelalte sunt facultative și conținutul lor trebuie să fie strict legat de ideea-cheie. „Fișele“ sunt separate între ele printr-un rând alb. Este suficient.

Atenție și la faptul că al treilea rând logic este reprezentat în pagină de două rânduri vizuale. Semnul „:“ ne arată limpede care sunt rândurile logice (doar cinci, în vreme ce sunt șase rânduri vizuale).

N-am mai complicat exemplul cu scrierea citatului în limba greacă

2. De la textul clasic la cel electronic

veche. Nu este un lucru pe care l-ar face frecvent cineva în activitatea curentă de management.⁹

Ce facem însă dacă vrem să „subliniem“ un cuvânt? Folosim un *marcaj* de genul `cunoașterea este _crezare_ adevărată și întemeiată`. Termenul *crezare* a fost scos în evidență prin punerea lui între spații subliniate. Pentru alte marcaje într-un fișier text vedeți schema standard de prezentare a unui caz.

„Fișele“ astfel construite nu trebuie confundate cu acelea scrise pe bucăți de hârtie de dimensiunile unei cărți poștale. De aceea am și vorbit despre „însemnări electronice“. Partea cea mai prețioasă la aceste însemnări este structura. Ceea ce am oferit aici este, desigur, un exemplu. Pot fi gândite și alte structuri și vom vedea pe parcurs soluții în acest sens.

Structura este prețioasă pentru că ușurează prelucrarea informației. Să zicem că vreau să aflu doar ce este relevant în însemnările pe care le consult. Voi examina doar rândurile care încep cu eticheta „cuvinte-cheie“.

Multe operații cu informațiile nu sunt foarte lesne de făcut folosind structura de mai sus. De exemplu, sortarea însemnărilor poate să fie mai dificil de operat. Nu tot ceea ce era facil de făcut cu fișele clasice este imediat ușor de realizat cu computerul.

2.2 Managementul și analiza cazurilor

Părintele managementului ca disciplină academică, Peter Drucker, vorbea despre cinci operații de bază în munca unui manager:¹⁰

- stabilirea obiectivelor;
- organizarea;
- crearea unei motivații la cei cu care lucrează;
- măsurarea și comunicarea rezultatelor;
- dezvoltarea persoanelor cu care lucrează, inclusiv a propriei persoane.

⁹Cine este curioasă sau curios să aplece la ce tehnică am recurs poate afla detaliile la adresa de web <<http://www.ub-filosofie.ro/~solcan/wt/gnu/a/agvim.html>>.

¹⁰Peter Drucker[1, p.400]

2. De la textul clasic la cel electronic

Care ar fi însă profilul muncii unui manager al informațiilor? Dacă ar fi să imităm schema propusă de Peter Drucker, atunci managerul informațiilor:¹¹

- stabilește obiectivele activităților legate de computerele firmei;
- organizează activitățile legate de computerele firmei;
- crearea unei motivații la cei cu care lucrează;
- măsurarea și comunicarea rezultatelor;
- formarea persoanelor cu care lucrează cu computere, inclusiv a propriei persoane.

Cum într-o firmă modernă computerele sunt foarte răspândite, „formarea persoanelor cu care lucrează cu computere“ privește multă lume. Chiar dacă nu aceasta este misiunea sa principală, într-o măsură mai mică sau mai mare, fiecare manager trebuie să cel puțin unele decizii legate de computere. Iar, de multe ori, decizii care par pur tehnice țin de fapt de managementul informațiilor.

Managementul nu este o disciplină apriorică, pur teoretică. Din contră, materialul empiric este extrem de important în management. Ca să desfășori munca de manager în sensul definit mai sus, ai nevoie să te sprijini pe un material empiric. Or, acest material nu se reduce la elemente pur cantitative. Nu poți face doar calcule numerice pentru a lua o decizie managerială. Aspectele calitative joacă un rol foarte important.

Cum să surprindem însă aspectele calitative? Cum să le prezentăm sistematic, în așa fel încât materialul empiric să fie relevant nu doar pentru o anumită situație, ci pentru un tip de situații? O tehnică larg folosită este cea a analizei de caz. În ciuda poate a numelui, o analiză de caz vizează ceea ce poate fi de folos, prin analogie, și-n alte cazuri.

În drept, sub o formă sau alta, jurisprudența joacă un rol foarte important. A analiza o speță, *mutatis mutandis*, înseamnă a analiza un caz. Analiza se face însă în așa fel încât sunt puse în evidență elemente care permit stabilirea de legături între cazuri. Doar în acest fel un caz poate fi de folos în practica judiciară.

Într-o analiză de caz, ca și atunci când ne facem o fișă de lectură, trebuie să conferim o structură uniformă „fișelor“. Dacă, de la un caz la altul, spunem „povestea“ în alt mod materialul empiric devine dificil

¹¹A se vedea și descrierea rolului managerului informațiilor pe situl Ministerului Muncii al SUA, la adresa <<http://www.bls.gov/oco/ocos258.htm>>.

2. De la textul clasic la cel electronic

de folosit. Nu putem căuta ușor printre cazuri. Nu le putem grupa și regrupa. Iar analogiile le vom face în mod arbitrar.

În figura „Analiza unui caz“ ilustrăm modul în care putem sistematiza un caz. Conținutul cazului este preluat, cu modificări, din cartea lui Rob Mattison despre bazele de date[5, pp.1-2]. Am prezentat însă cazul urmând o structură. Ca și-n cazul fișei de lectură, structura are „câmpuri“. În fișa tradițională a unui caz, evidențiem aceste câmpuri vizual.

Problema	Ion este anticar. El are un mic magazin într-un mare complex comercial. S-au dus însă vremurile când un anticar ținea minte tot ce are în magazin. Ion s-a hotărât să-și ia o bază de date.
Soluții alternative	Soluțiile comerciale care i se oferă merg de la 100RON la 3000RON. Soluțiile cu sursă deschisă sunt de la 0RON la 120RON.
Operații tehnice	<ul style="list-style-type: none">• cumpărarea unui computer;• instalarea bazei de date;• învățarea utilizării bazei de date;• proiectarea sistemului anticarului;• scrierea sistemului (programului).
Decizii manageriale	Ion este un om energic și ar putea învăța să scrie și programul bazei de date. Probabil va angaja însă un consultant care va scrie programul de care are nevoie pentru a utiliza baza de date.
Costuri și beneficii	Dacă Ion întreprinde singur toate acțiunile va consuma bani și timp. Trebuie văzut dacă angajarea unui consultant nu face, de fapt, să scadă costurile.

Adaptare după Rob Mattison

Figura 2.2: Cazul anticarului

Cum am putea proceda însă, tot prin analogie cu însemnările de lectură, pentru a „fișa“ un caz? De această dată este vorba despre un text mai lung. Am putea să recurgem tot la text pur? De ce nu? Examinați sursa unei pagini *wiki*¹² sau un text din Biblioteca Gutenberg¹³. Paginile *wiki* sau cărțile Gutenberg în format *plain text* folosesc marcaje pentru a indica structura ierarhică a textului sau pentru a evidenția porțiuni de text.

O formă simplificată de marcaj este propusă și aici, pentru analizele de caz. În esență, în textul de prezentare al unui caz, folosim marcaje

¹²Populara enciclopedie Wikipedia folosește acest format. Pentru lista enciclopediilor vedeți <http://meta.wikimedia.org/wiki/List_of_Wikipedias>. La articole, dați clic pe „modifică pagina“ și veți vedea cum este organizat logic textul, nu vizual.

¹³Pe Internet la <<http://www.gutenberg.org/>>. Uitați-vă la cărțile în format „plain text“.

2. De la textul clasic la cel electronic

pentru a preciza segmentele structurale ale textului și pentru sublinieri. Este deci o formă foarte simplă de a pune marcaje în text.

Ca și-n cazul însemnărilor de lectură, vom introduce practic marcajele, reluând exemplul din figura „Cazul anticarului“. Pentru a nu lungi însă textul folosim puncte de întrerupere.

__Cazul anticarului__

Ion este anticar.
Magazinul său este ...

==Problema==

S-au dus vremurile ...
Ion vrea o bază de date ...

==Soluții alternative==

Soluții "cu sursă închisă" ...
Prețul lor ...

Soluții "cu sursă deschisă"

===Operații tehnice===

====Achiziția computerului====

Cumpărarea computerului este primul pas.
Prețul ...

====Baza de date====

Instalarea bazei de date ...

Proiectarea sistemului ...

====Utilizarea bazei de date====

Învățarea utilizării bazei de date presupune ...

Rolul bazei de date în magazin este ...

===Decizii manageriale===

2. De la textul clasic la cel electronic

Acesta este un caz clasic de
produc sau cumpăr

Un consultant ar putea ...

==Avantaje și dezavantaje==

A _produce_ singur ...

Costurile implicate ...

Avantajele _cumpărării_ serviciilor ...

--Surse:--

Cazul este adaptat după ...

A se vedea și situl:

<<http://un-site-ipotetic/...>>

Fiecare caz este normal să aibă un titlu, pentru a ne putea referi ușor la el. Titlul l-am marcat cu `--` în vreme ce o subliniere normală din text este indicată prin `_`. Sursele trebuie și ele indicate. Le-am marcat cu `--` :`--` pentru a nu se confunda cu titlul. Adresa de Internet este pusă în paranteze unghiulare.

Momentele obligatorii ale analizei cazului sunt: problema, soluțiile alternative și precizarea avantajelor și dezavantajelor deciziei luate în cazul respectiv. Acestea sunt plasate la un prim nivel de secționare, indicat prin `==` `==`. Se observă cum nivelurile de secționare ulterioară a textului sunt indicate prin sporirea cu unu a semnelor egal, în dreapta și-n stânga.

Fiecare propoziție este scrisă pe alt rând logic. Alineatele sunt separate între ele cu un rând alb. Gruparea pe alinee trebuie și ea să reflecte „logica textului“. Propozițiile alineatului gravitează în jurul unei singure idei-cheie pentru alineatul respectiv. Trecerea la o altă idee-cheie impune construirea unui nou alineat. Deci rândul alb are și el rolul său; face parte din sistemul de marcaje introduse în text.

De ce am scrie însă atât de „inestetic“? Nu rezolvă aranjarea în pagină, tipul de literă ales mai „frumos“ decât marcajele acestea problema structurării textului prezentării? Categorie nu? Scopul este punerea în evidență a structurii logice a textului. Vizualul ar distruge această structură. Diferite editoare WYSIWG vor pune marcaje (ascunse) diferite. Va fi imposibil să menținem apoi continuitatea fluxului informațiilor.

De asemenea, trebuie să ne gândim că este vorba despre un text pe care-l vor citi (și) programele de calculator. Un om își poate da

2. De la textul clasic la cel electronic

seama care este titlul textului și din conținutul acestuia. Programului îi trebuie un marcaj special.

Rolul vizualului în exemplul de mai sus nu este totuși nul. Titlurile secțiunilor le-am pus pe centru. Sursa am alineat-o către dreapta. Aceste elemente nu au însă decât rolul de a ușura parcurgerea textului de către om. Ca atare ele nu joacă rolul unor marcaje. Un ipotetic program de computer care ar citi textul „Cazul anticarului“ ar șterge de la bun început spațiile albe sau „tab“-urile folosite pentru a centra textul. De asemenea, ar fi recomandabil ca, în ipoteza atașării fișierului respectiv la un e-mail, aceste elemente să fie eliminate, păstrându-se doar rânduri de text care încep fără spații albe. Programul de vizualizare a e-mail-ului va da peste cap aproape sigur toate centrările și alinierea la dreapta ale textului.

Adevărul este că nici pentru om sistemul de marcaje puse explicit în text nu este un lucru rău. El ne obligă să respectăm un standard în prezentarea cazurilor. Constrângerile dictate de tehnologia informației ne forțează să ne ordonăm gândurile. Asta nu poate fi deloc un lucru rău pentru decizia managerială.

2.3 De la textul clasic la cel electronic: termeni-cheie

- fișă de lectură
- analiză de caz
- separator (al informațiilor din text)
- text cu marcaje

2.4 De la textul clasic la cel electronic: exerciții și analize de caz

Exercițiul 1: Dacă nu folosiți cumva Unix, instalați editorul de texte Vim pe computerul dumneavoastră. Căutați și alte editoare de text pur. Atenție la faptul că acestea sunt „programmer’s editor”, deci le veți găsi pe Internet ca editoare pentru scrierea (în special) a programelor de calculator.

Exercițiul 2: Studiați, cu ajutorul Vim, deosebirea dintre rândurile logice și cele vizuale.

2. De la textul clasic la cel electronic

Exercițiul 3: Respectând formatul indicat în această unitate de învățare citiți toate unitățile de învățare dedicate managementului informației și elaborați un fișier de tip text cu însemnări electronice de lectură. Folosiți Notepad, Vim sau alt editor de acest tip. Nu folosiți în nici un caz un editor pentru tehnoredactare de tip WYSIWG.

Exercițiul 4: Respectând formatul indicat în această unitate de învățare, pe baza unor surse identificate în mod independent de către dumneavoastră, prezentați analiza unui caz. Respectați sistemul de marcaje indicat. Folosiți Notepad, Vim sau alt editor de acest tip. Nu folosiți în nici un caz un editor pentru tehnoredactare de tip WYSIWG.

Exercițiul 5: Procurați-vă de pe Internet sau din alte surse legal accesibile fișiere de tip text produse sub platforme diferite (Windows, Mac sau Unix). Examinați diferențele de codificare a sfârșitului de rând logic.

Exercițiul 6: Atașați la e-mail-uri trimise către propria dumneavoastră adresă textele produse la exercițiile 3 și 4. Studiați rezultatul. Faceți o comparație cu „attachment“-uri care nu sunt în format text.

Exercițiul 7: Dacă aveți posibilitatea, studiați modul în care pot fi citite fișierele atașate la e-mail în condițiile folosirii unor programe diferite de citire a e-mail-ului. Studiați diferența dintre programele de citire în mod text a e-mail-ului și cele în mod grafic. Dacă puteți, vedeți și care sunt diferențele dintre platforme.

Analiza de caz 1: O firmă achiziționează o bază de date scumpă pentru a evidențierele sale electronice. În practică instalează sistemul doar pe un computer aflat undeva într-un birou. Restul computerelor erau prea vechi pentru a instala pe ele noul sistem. Apoi nu se concepuse nimic pentru a transfera automat datele de pe vechiul sistem pe noul sistem. Articole de inventar din depozit erau câteva zeci de mii. Sistemul achiziționat va fi avut sens la sute sau poate milioane de articole de inventar. Managerii care au luat deciziile n-aveau efectiv idee despre ceea ce se întâmplă într-un sistem informatic. Analizați cazul și sistematizați prezentarea după schema indicată în această unitate de învățare. Evaluați decizia luată de către managerii firmei. Evaluați importanța cunoștințelor de tehnologia și de managementul informațiilor pentru decizia luată.

Analiza de caz 2: Un manager are de ales între folosirea unor fișiere al căror format este public (se știe de către toată lumea cum sunt structurate fișierele respective) și fișiere al căror format nu a fost făcut public. Analizați cazul. Căutați eventual exemple concrete în practica dumneavoastră.

UNITATEA DE ÎNVĂȚARE 3

Extragerea informațiilor

Este o mare deosebire între a încerca să dobândești cunoaștere și a porni în căutarea unor informații. Nu există o algoritmică în căutarea cunoașterii. În schimb, căutarea informațiilor se face sub semnul algoritmicii.

Căutarea informațiilor poate să fie mai ușoară sau mai dificilă în funcție de felul în care au fost stocate informațiile. Structura lor joacă aici un rol aparte. Legăturile deja create între informații facilitează căutarea.

Ca și-n restul manualului vom privilegia informația conținută în texte. Așadar prin „căutarea informațiilor“ trebuie înțeles procesul de redescoperire a informațiilor stocate în texte.

Obiectivele acestei unități de învățare sunt:

- Înțelegerea distincției dintre introducerea unor simple opțiuni într-o interfață pentru căutarea într-un text și folosirea unui limbaj de căutare;
- Însușirea unor elemente de bază ale prelucrării textului cu ajutorul substituițiilor;
- Înțelegerea tehnicilor de bază ale sortării;
- Precizarea modului în care un fișier text poate fi folosit ca bază de date;
- Ilustrarea modului în care tehnicile introduse anterior pot fi folosite integrat pentru a construi și consulta dicționare.

Pentru a aprecia miza obiectivelor acestei unități de învățare cel mai bun lucru ar fi să începeți prin a citi articolul indicat în exercițiul final. Încercați să desprindeți de acolo ideea de extragere a informațiilor (în

3. Extragerea informațiilor

cazul discutat în articolul respectiv, dintr-o pagină de *web*). Evident, nu încercați să puneți direct în aplicare cele citite în articol. Reveniți la prima secțiune a unității de învățare. Dacă veți citi rezolvând în același timp și exercițiile nu vă va veni greu probabil să înțelegeți și chiar să testați programul demonstrativ din ultimul exercițiu.

3.1 Căutarea informațiilor

Există un punct de întâlnire între logică și managementul informațiilor care este foarte important. În logică, studiem procesele prin care se trag concluzii adevărate din premise adevărate. Nu toate propozițiile sunt de la bun început asumate; inferențele sunt cele care ne permit să extindem treptat sfera propozițiilor acceptate. Prin analogie, putem vorbi de un proces prin care informațiile sunt făcute explicite sau sunt extrase.

În „Cazul anticarului“ ne-am referit la alternativa dintre *a produce și a cumpăra*. A produce structurile specifice bazei de date a anticariatului este mai degrabă un serviciu pe care-l poate furniza un specialist. Ar fi însă greu de crezut că vom apela la un consultant plătit ori de câte ori căutăm cuvinte într-un text. Plătim însă, sub o formă sau alta, pentru unealta cu care căutăm.

Vizualizoarele de pagini de *web*, editoarele de texte, editoarele pentru tehnoredactare și multe alte programe au incluse unelte de căutare. Vom da, mai greu sau mai ușor, peste „*find*“. Este puțin probabil că vom scrie propriul nostru „*find*“. Vom folosi unul dintre cele existente.

Voi utiliza aici - ca exemplu - editorul Vim. Vim are specificul său: este un editor *modal*. Alegerea între editoare modale și non-modale constituie un veritabil caz. Pentru a sublinia părțile aflate „în conflict“ vom da un nume cazului după modelul din dreptul jurisprudențial:

__Editor modal vs. editor non-modal__

==Problema==

Scrierea unui text cu ajutorul computerului presupune nu doar introducerea de litere, ci și comenzi.

De pildă, trebuie să căutăm cuvinte.

Ar fi imposibil însă să rezervăm o tastă pentru căutarea de cuvinte ș.a.m.d.

Tastatura ar căpăta dimensiuni prea mari.

3. Extragerea informațiilor

[¹ Așa cum sugerează și sintagma „editor modal“, aceste editoare au moduri. În funcție de mod, apăsarea unei taste produce efecte diferite. La Vim, modul normal este cel al comenzilor! În modul normal, apăsarea tastei `j` produce deplasarea cursorului în jos. Apăsarea tastei `i` ne aduce în modul *insert*. În acest mod, tastele inserează semne în text. Într-un editor non-modal se folosesc pentru comenzi combinații de taste. De asemenea, se pot folosi meniuri. Un avantaj al Vim este acela că poți tasta orice - litere sau comenzi - fără a deplasa prea mult degetele pe tastatură. Vim este perfect compatibil cu „metoda oarbă“ de tastare.]

Căutarea într-un text de mici dimensiuni nu pune totdeauna în evidență avantajele uneltelor din computer. Lucrurile se schimbă când este vorba despre o carte, cum ar fi de pildă lucrarea de pionierat în domeniul managementului a lui Fredrick Winslow Taylor, *The Principles of Scientific Management* (1911).

Cel mai simplu lucru este desigur să căutăm un cuvânt în text. Pentru aceasta aducem Vim în modul normal. Acesta este modul în care este Vim la pornire, dar pentru a fi siguri apăsăm `<Esc>`. Dacă, în modul normal apăsăm tasta `g` de două ori, cursorul se va duce la începutul textului. De acolo, când apăsăm tasta cu bară oblică `/` cursorul se mută pe ultima linie și putem da tasta cuvântul pe care vrem să-l căutăm. În cazul cuvântului „management“ șirul de comenzi poate fi scris pe scurt astfel:

`<Esc>gg/management<CR>`

Rezultatul este cel din figura „Căutarea termenului «management»“. Nu trebuie să vă faceți griji dacă aspectul Vim este ușor diferit pe computerul dumneavoastră. Vim este un editor programabil, extrem de flexibil, iar fiecare și-l configurează în funcție de propriile dorințe. Ceea ce ar trebui să apară după apăsarea pe *Enter*, `<CR>`, în limbajul Vim este însă o schimbare a fondului pe care este scris termenul „management“. Dacă ea nu apare, trebuie verificată configurarea pe care o are Vim.

Un alt exercițiu simplu este căutarea sfârșitului rândurilor logice. În cazul Vim se rezolvă folosind comenzile: `<Esc>gg/$<CR>` Căutarea în textul cărții lui Taylor ne rezervă o surpriză, nu tocmai plăcută. Rândurile vizuale ale cărții sunt realizate prin apăsări pe *Enter*. Cu alte cuvinte rândurile logice servesc la realizarea rândurilor vizuale. Vom vedea destul de repede care este dezavantajul atunci când vrem nu să citim cartea pe ecran, ci să căutăm în text.

¹Pentru a economisi spațiu transcrierea cazurilor sub o formă standard este lăsată ca exercițiu. V. secțiunea „Evaluare“ din cadrul unităților de învățare.

3. Extragerea informațiilor

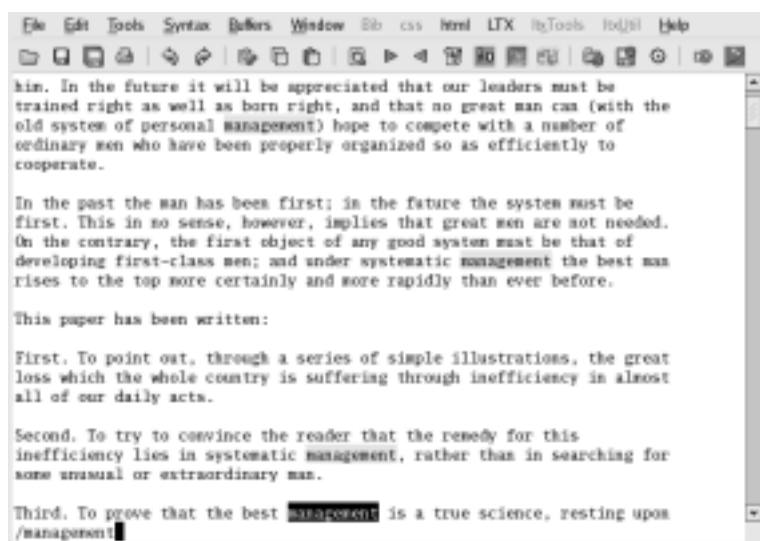


Figura 3.1: Căutarea termenului „management“

Un verb precum „a extrage“ sugerează imediat alte metafore: a fora, a scormoni, a sonda. Dacă vrem, de pildă, să vedem toate rândurile pe care apare „science“ în textul lui Taylor? Atunci unealta potrivită se numește „grep“.

Taylor are o predilecție pentru „*science*“. Este și specificul acelor vremuri, când totul se vrea „*scientific*“. *Grep* poate identifica toate rândurile în care apare șirul de semne „*science*“. ² Faptul că extrage rândurile în care apare șirul de semne căutat este un avantaj: în acest fel avem și contextul în care apare respectivul șir.

În sistemele Unix *grep* este apelat din „linia de comandă“. Asta nu înseamnă altceva decât că trebuie să deschidem un terminal, similar cumva cu micile ecrane negre din Windows sau cu ecranul vechiului MSDOS. Comenzile sunt tastate în fereastra respectivă.

Pentru a obține începutul fișierului care se vede în figura „Rânduri cu șirul de semne «*science*»“ am folosit comanda:

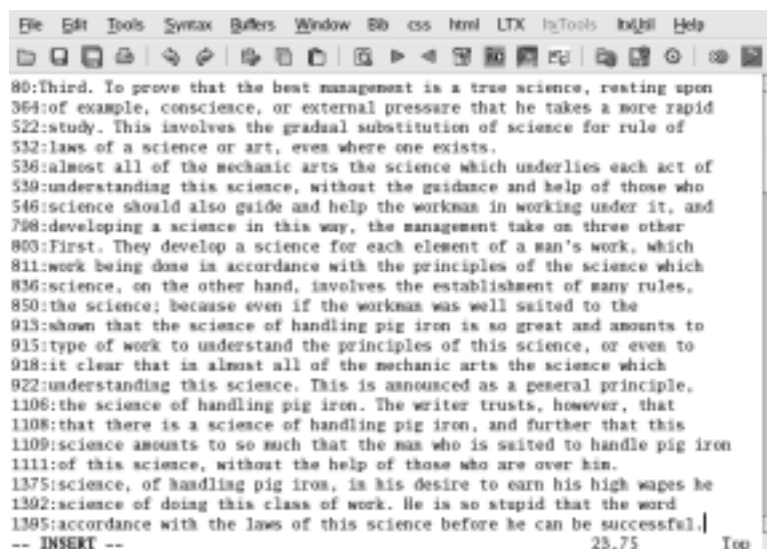
```
grep -n science pscmg10.txt > pscmg10-science.txt
```

Numele fișierului cu textul cărții lui Taylor, preluat din Biblioteca Gutenberg este „pscmg10.txt“. Acel `-n` de după numele uneltei *grep* îi spune programului respectiv să afișeze numărul rândului. Partea finală a întregii comenzi, `> pscmg10-science.txt`, cere ca rezultatul să fie

²„Șirul de semne“, pentru că „*science*“ poate să un cuvânt sau o parte a unui cuvânt, cum ar fi „*conscience*“. V. și mai jos.

3. Extragerea informațiilor

transferat într-un fișier. Altfel, el este afișat în terminal (în fereastra unde este tastată comanda).



```
File Edit Tools Syntax Buffers Window SB css html LTX InTools InDiff Help
80:Third. To prove that the best management is a true science, resting upon
364:of example, conscience, or external pressure that he takes a more rapid
522:study. This involves the gradual substitution of science for rule of
532:laws of a science or art, even where one exists.
536:almost all of the mechanic arts the science which underlies each act of
539:understanding this science, without the guidance and help of those who
546:science should also guide and help the workman in working under it, and
798:developing a science in this way, the management take on three other
803:First. They develop a science for each element of a man's work, which
811:work being done in accordance with the principles of the science which
836:science, on the other hand, involves the establishment of many rules.
850:the science; because even if the workman was well suited to the
913:shown that the science of handling pig iron is so great and amounts to
915:type of work to understand the principles of this science, or even to
918:it clear that in almost all of the mechanic arts the science which
922:understanding this science. This is announced as a general principle.
1108:the science of handling pig iron. The writer trusts, however, that
1108:that there is a science of handling pig iron, and further that this
1109:science amounts to so much that the man who is suited to handle pig iron
1111:of this science, without the help of those who are over him.
1375:science, of handling pig iron, in his desire to earn his high wages he
1392:science of doing this class of work. He is so stupid that the word
1395:accordance with the laws of this science before he can be successful.
-- INSERT -- 25,75 Top
```

Figura 3.2: Rânduri cu șirul de semne „science“

Acum se vede și care este dezavantajul folosirii rândurilor logice pentru a realiza rânduri vizuale. Avem un context pentru „*science*“, dar acesta nu este, de regulă, format din propoziții întregi.

Cu ajutorul *grep* putem începe și operațiile de realizare a unor statistici. Cum Vim indică automat numărul rândurilor din fișier, putem afla că sunt 72 de rânduri (vizuale !) în care Taylor folosește șirul de semne „*science*“. Având la dispoziție și numerele de rând, ne putem face o idee și despre distribuția acestor apeluri la ideea de știință.

Un alt avantaj al *grep* este că putem căuta cu ajutorul său în mai multe fișiere deodată, inclusiv „în adâncime“ (în folderurile aflate în folderul de la nivelul căruia am început căutarea).

Taylor n-avea cum să știe că pe noi ne-ar interesa de câte ori și în ce context invocă el ideea de știință. Informația este însă implicit prezentă în text. Totul e să găsim un limbaj adecvat pentru a extrage cu precizie³ informația pe care o căutăm.

³Separând, de pildă, pe „*science*“ de „*conscience*“, găsim un context convenabil ș.a.m.d.

3.2 Limbajul căutării informațiilor

Adevărata putere a căutărilor din Vim și a grep este dată de *expresiile regulate*. Sintagma „expresii regulate“ desemnează de fapt un limbaj: un limbaj folosit pentru a căuta în text. Faptul că avem de a face cu un limbaj ne pune în fața unui caz deosebit de interesant:

```
__Limbaj vs. interfață grafică__  
  
==Problema==
```

Puterea computerului este dată de limbajele pe care le „înțelege“. Este posibil însă să folosim computerul fără a ști vreunul dintre aceste limbaje, recurgând la interfețe grafice pe care, în esență, bifăm diverse opțiuni. Problema este însă să determinăm în ce condiții este mai bine să recurgem la expresii în limbaje, pe care le înțeleg computerele și în ce condiții sunt preferabile interfețele grafice.

[Instanța cea mai cunoscută a acestui caz este probabil „CLI vs. GUI“. „CLI“ este *comanda în linie* (ca la folosirea *grep* într-un terminal Unix). „GUI“ este *Graphic User Interface*, care transformă computerul într-un soi de tablou de bord pe ale cărui butoane apăsăm pentru a da comenzi. Succesul unor programe, precum editorul pentru tehnoredactare „Word“, se explică în bună parte prin interfața grafică reușită. Pe măsură ce interfața grafică a unui program încearcă să prindă însă tot mai mult din puterea unui limbaj de computer, interfața devine tot mai plină de meniuri, butoane și opțiuni, care fac operarea dificilă. Decizia în favoarea unei soluții sau alta ține în parte de tehnologie și în parte de management, de luarea în considerare a ceea ce știu oamenii și a modului în care pot fi mai productivi.]

Limbajul expresiilor regulate extinde considerabil posibilitățile de a lucra cu Vim, grep sau alte programe. În același timp, dificultatea utilizării sale este recunoscută chiar și de programatoarele sau programatorii cu experiență. De aceea ne vom limita aici la câteva utilizări elementare ale expresiilor regulate.

În textul său, Taylor pare să recurgă la enumerări de genul „*First, Second, Third* ș.a.m.d.“. Cât de des? În ce context? În cazul termenului „*First*“ ar avea sens să-l căutăm pur și simplu cu *grep* prin tot textul. Să zicem însă că vrem să găsim „*First*“ doar la început de rând. Comanda este:

3. Extragerea informațiilor

```
grep -n "^First" pscmg10.txt
```

Expresia regulată este scrisă între ghilimele. Faptul că am pus `"^ "` nu este decât un mod de a spune că expresia care urmează trebuie să apară la început de rând.

Expresiile regulate pot remedia o problemă pe care o generează, de pildă, căutarea rândurilor care cuprind cuvântul „*science*”: rândurile care conțin „*conscience*” apar și ele printre rezultate! Problema dispăre dacă scriem în linia de comandă:

```
grep -n "\<science\>" pscmg10.txt > pscmg10-science.txt
```

Comanda de mai sus conține expresia `\<science\>` care nu este altceva decât un fel de a spune că ne interesează doar cuvântul întreg, nu și „*science*” ca parte a unui alt cuvânt. Firește că și numărul de rânduri extrase se modifică.⁴ Pentru a conferi un rol special parantezelor unghiulare se recurge la barele oblice inverse.

Limbaajul devine cu adevărat interesant însă când folosim expresii precum cea din comanda de mai jos:

```
grep -n "\<science\>\\|scientific" pscmg10.txt
```

În acest caz, operatorul `\\|` îi spune programului să caute rânduri pe care se află `\<science\>` sau `scientific`.

Dacă vrem să găsim rânduri pe care se găsesc atât „*science*”, cât și „*scientific*”, putem folosi comanda:

```
grep -n "science.*scientific" pscmg10.txt
```

În expresia `science.*scientific` punctul ține locul oricărui semn, iar asteriscul arată că aceste semne oarecari (deci nu neapărat același semn) se repetă de un număr de ori.

Expresiile regulate pot fi folosite și în căutările Vim. De fapt, când am căutat sfârșiturile de rând cu Vim am folosit o expresie regulată foarte simplă. Semnul dolarului îi „spune” programului că avem în vedere sfârșitul de rând (evident, în sens logic).

Limbaajul expresiilor regulate are „dialecte”. Expresiile care merg cu *grep* s-ar putea să trebuiască să fie modificate în Vim și invers. De asemenea, Vim poate folosi mai multe dialecte ale limbaajului expresiilor regulate.⁵

⁴Nu sensibil în cazul textului lui Taylor, care folosește termenul „*conscience*” doar o singură dată.

⁵Consultați documentația Vim pentru a vedea despre ce este vorba și cum puteți comuta între variante diferite ale expresiilor regulate.

3. Extragerea informațiilor

Expresiile regulate pot fi folosite și pentru scopuri neonorabile. Este relativ ușor să fie utilizate pentru a extrage adrese de e-mail din pagini de web, în scopul trimiterii de *spam*, după cum se vede din cazul următor:

`--Adrese de e-mail în pagini de web și spam--`

`==Problema==`

„Spam“ sunt mesajele e-mail „deversate“ către adrese la care n-au fost solicitate. Cei care trimit spam trebui să afle însă adresele.

Una din sursele din care se pot afla adrese de mail sunt paginile de web pe care adresele apar sub forma: `nume@furnizor-de-e-mail.com` sau ceva similar.

Prezența unei astfel de adrese este extrem de ușor de detectat de programe care caută prin paginile de web folosind expresii regulate adecvate.

[Există diverse soluții în acest caz. Un lucru este cert: afișarea „în clar“ (ca text) a adresei de e-mail în pagina de web este calea cea mai sigură către un val de *spam*.]

3.3 Caută și înlocuiește

Cei mai mulți oameni sunt obișnuiți cu substituțiile operate în text cu ajutorul unui editor de tehnoredactare. Modurile de a substitui în fișierele text sunt însă mult mai variate, subtile și, nu o dată, înșelătoare.

Vom reveni mai întâi la Vim și la textul cărții lui Taylor. Ce am putea face pentru a pune fiecare propoziție pe un rând logic? Cum fișierul descărcat de pe Internet are peste 3500 de rânduri am consuma prea mult timp dacă am schimba manual sfârșiturile de rând.

Trebuie să efectuăm niște substituții. Vim ne dă posibilitatea de a combina într-o singură comandă căutarea și substituirea într-un șir de semne de pe un rând. Pentru aceasta treceti în modul normal și apoi apăsați „două puncte“. Formal acest lucru s-ar scrie `<Esc>:`. După ce ați apăsat „două puncte“ veți vedea cum cursorul se deplasează în partea de jos a ecranului și Vim așteaptă o comandă. Vim este laconic, așa că `:s/Mr\.` `Roosevelt/Mr.Roosevelt/` spune o mulțime de lucruri. Litera `s` este o comandă (substituie!). Cele două puncte *nu* trebuie

3. Extragerea informațiilor

tastate din nou. Ele indică doar faptul că Vim a fost trecut în modul comandă în linie. Între prima și a doua bară oblică se află o expresie regulată care trebuie căutată. Între a doua și a treia bară oblică se află expresia cu care se substituie. Evident, pentru a executa comanda trebuie să apăsam *Enter*.

Studiind figura „Eliminarea unui spațiu alb“ ne putem face o idee despre modul în care funcționează o comandă de substituie. Atenție! Așa cum a fost ea dată, comanda se aplică doar pe rândul pe care se află cursorul și doar o singură dată.

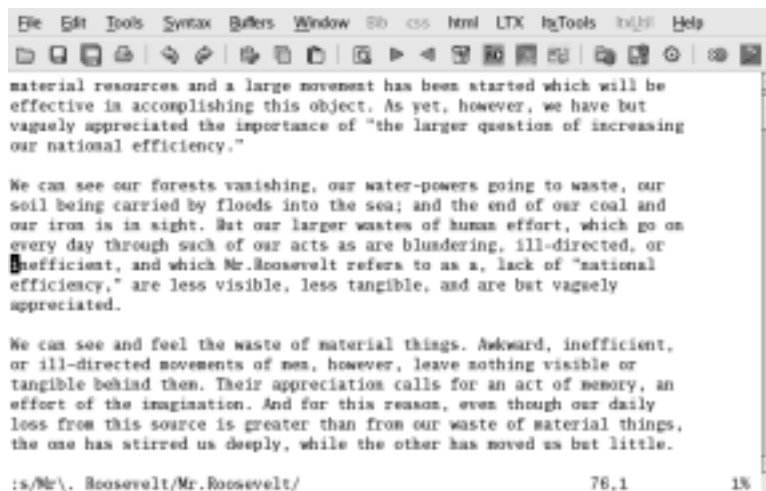


Figura 3.3: Eliminarea unui spațiu alb

Cu titlu de exercițiu, putem încerca să punem pe fiecare rând logic o propoziție. Procedura descrisă aici nu dă un rezultat perfect, dar ne arată cam ce putem face pentru a modifica modul în care sunt plasate sfârșiturile de rând.

Mai întâi „lipim“ rânduri care se termină cu un cuvânt, o virgulă sau ghilimele. Sub Linux, putem încerca următoarea comandă:

```
:%s/\([a-z '][a-z,"]\)\n/\1 /
```

Semnele `\n` arată că sistemul trebuie să se uite la sfârșiturile de rând. Între parantezele rotunde precedate de bare oblice inverse se află ceea ce-n limbaajul expresiilor regulate se numește un „grup“. Vim identifică și reține în memorie orice grup format din două semne aflate la sfârșitul rândului. Spectrul acestor semne este indicat între parantezele drepte. Penultimul semn pe rând poate să fie o literă mică, un spațiu alb sau un apostrof. Expresia pentru ultimul semn de pe rând poate fi înțeleasă prin analogie cu aceea pentru penultimul semn. Substituția

3. Extragerea informațiilor

este foarte simplă: `/\1 /` arată că este pus la loc grupul marcat de parantezele rotunde precedate de bare oblice inverse plus un spațiu alb. Omitem deci sfârșitul de rând. Spațiul alb îl punem pentru a nu lipi între ele cuvinte. În sfârșit, semnul procentului pus înaintea comenzii de substituie este foarte important: el cere sistemului să opereze substituția pe toate rândurile din fișier.

Rezultatul obținut este doar unul cu caracter intermediar. În cazul textului cărții lui Taylor, vor fi puse practic pe un rând logic aliniate întregi. Pentru a progresa mai facem o schimbare cu ajutorul acestei comenzi:

```
:%s/\([a-z]\.\)[ ]+\([A-Z]\)/\1\r\2/g
```

De data aceasta căutăm două grupuri. Primul se termină cu un punct urmat de unul sau mai multe spații albe. Al doilea grup începe cu majusculă. La substituie introducem între grupuri un sfârșit de rând. Atenție la faptul că sfârșitul de rând diferă de la o platformă la alta! Comanda aceasta am folosit-o sub Unix.

După transformările suferite, o căutare a cuvântului „*science*” îmi oferă un context mai bogat.

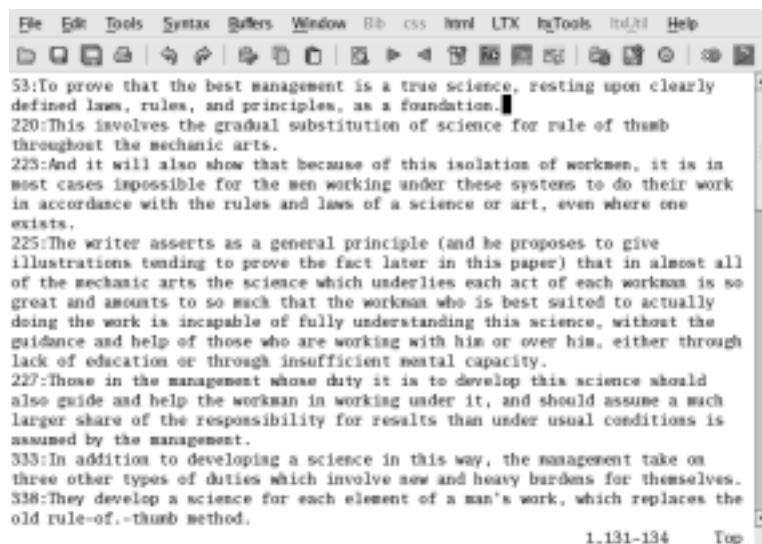


Figura 3.4: Lărgirea contextului căutării

Ne putem face o idee și despre felul în care arată fișierul după rearanjarea cu ajutorul substituțiilor din figura „Propoziții pe rânduri logice”

Este posibil să operăm substituții în fișiere text „din mers”, „pe flux”, fără a deschide fișierul într-un editor precum Vim. Pentru aceasta avem nevoie însă de un editor precum *sed*, aflat îndeobște pe sistemele de tip Unix.

3. Extragerea informațiilor

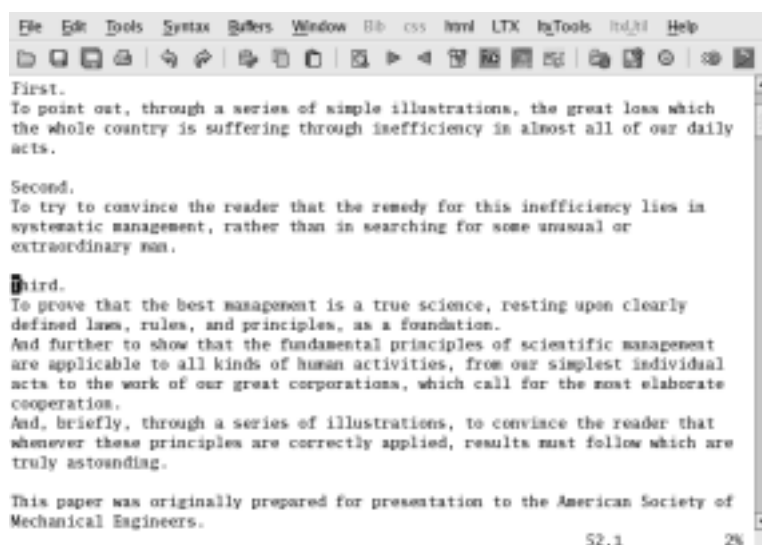


Figura 3.5: Propoziții pe rânduri logice

3.4 Sortarea informațiilor extrase

După ce am extras rânduri dintr-un text, am dori desigur, de multe ori să le sortăm. Sunt și multe alte situații în care vrem să sortăm rândurile unui fișier de tip text.

În exemplul care urmează nu conținutul informațiilor contează, ci modul în care sunt manevrate. Am creat un mic fișier de tip text în care am trecut pe fiecare dintre cele trei rânduri o dată calendaristică⁶ și valoarea unei unități de investiții la un ipotetic fond mutual.⁷ Iată ce conține acest fișier:

1/31/69	40
12/31/76	300
12/31/84	3000

În Unix există o unealtă *sort* care ordonează rândurile dintr-un fișier text. Chiar dacă sunteți în Windows v-aș sfătui să eliminați programul *sort*, moștenit din vechiul MSDOS și să folosiți versiunea Unix a lui *sort*.

Să vedem ce rezultat produce apelarea lui *sort*:

⁶Puteți afla mai multe despre problema reprezentării datelor calendaristice pornind de la articolul din Wikipedia despre standardul ISO în materie, „ISO 8601“ <http://en.wikipedia.org/wiki/ISO_8601>.

⁷Exemplul este inspirat în mod liber din tabelul de la pagina 146 a cărții lui George Soros[12]. Valoarea unității de investiții este în sensul NAVPS (v. <<http://www.investopedia.com/terms/n/navpershare.asp>>.

3. Extragerea informațiilor

```
sort curs1.txt
12/31/76      300
12/31/84      3000
1/31/69       40
```

Datele calendaristice au fost sortate ca și cum ar fi un text! Operez câteva modificări în fișier și-l aduc la forma:

```
12/31/76      300
01/31/69       40
15/31/84      3000
```

Sortarea se va face acum altfel. Firește, se va face după luni!

```
01/31/69       40
05/31/84      3000
12/31/76       300
```

Folosind subtilitățile programului *sort* din Unix pot sorta după ani.

```
sort -t/ -k3,3 curs2.txt
01/31/69       40
12/31/76       300
05/31/84      3000
```

De data aceasta fișierul este văzut ca având pe fiecare rând câmpuri. Bara oblică separă câmpurile. Semnul care urmează după opțiunea *-t* îi spune programului de sortare ce separator să folosească. După opțiunea *-k* urmează numerele care arată câmpurile după care se sortează. Atenție însă la faptul că ultimul câmp cuprinde atât anul, cât și valoarea unității de investiții!

Având în vedere problemele ridicate de datele calendaristice cred că este mai prudent să folosim date calendaristice de forma „an-lună-zi”. Iată și a treia formă a micului meu fișier:

```
76-12-31      300
69-01-31       40
84-05-31      3000
```

Chiar și-n această formă datele calendaristice pot fi problematice. Nu vă amintiți de problema anului 2000? Dacă rezultatele fondului sunt gândite pe termen lung? Ar putea fi vorba despre ceva de genul:

3. Extragerea informațiilor

1976-12-31	300
2069-01-31	40
3084-05-31	3000

Folosirea formei complete a anului este evident dezirabilă în cazuri de acest tip.

Cum s-ar face sortarea după valoarea unităților de investiții? Voi construi o a cincea formă a fișierului. În această formă separ două câmpuri în felul următor:

1969-01-31	:	0400
1976-12-31	:	0030
2004-05-31	:	3000

Cele două puncte de pe fiecare rând servesc drept separator. Pe primul câmp apar date calendaristice (într-o formă - sperăm - ușor de sortat) și pe al doilea câmp apar valori ale unității de investiții. M-am gândit că al doilea câmp este văzut ca un text și am pus zerouri pentru ca sortarea să aibă loc corect:

1976-12-31	:	0030
1969-01-31	:	0400
2004-05-31	:	3000

Valoarea acțiunilor ipoteticului fond este cea mai scăzută la 31 decembrie 1976!

Adevărul este că *sort* din Unix poate recunoaște numere. Putem sorta un fișier de genul:

1969-01-31	:	400
1976-12-31	:	30
2004-05-31	:	3000

Procedăm în felul următor:

```
sort -t: -k2,2 -n -b curs6.txt
1976-12-31 : 30
1969-01-31 : 400
2004-05-31 : 3000
```

Opțiunea *-n* îi spune programului *sort* să considere că are de face cu numere. Opțiunea *-b* cere ignorarea spațiilor albe. S-ar putea ca ea

3. Extragerea informațiilor

să fie activată automat. Oricum, dacă o puneți, acest lucru n-are de ce să dăuneze.

Capcane chiar mai complicate întinde sortarea în momentul în care am lucra cu un fișier cu nume de familie românești. Să zicem că este vorba de data nașterii unor persoane:

```
Stere      : 1987-10-23
Tomescu    : 1979-03-15
Ștefan     : 1984-07-04
```

Sortarea lor după data nașterii nu prezintă nici o problemă. Tomescu este, evident, cel mai în vârstă. Putem sorta ușor și-n ordine inversă:

```
sort -t: -k2,2 -b -r data-nasterii.txt
Stere      : 1987-10-23
Ștefan     : 1984-07-04
Tomescu    : 1979-03-15
```

Pentru a sorta în ordine inversă folosim opțiunea `-r`.

Sortarea după nume este însă problematică. În calculator literele au asociate coduri. Programul `sort` se folosește, de fapt, de aceste coduri pentru a sorta. În fond trebuie să sorteze după un criteriu! Problema este că *S* are, de pildă, codul *83* și *T* are codul *84*, iar *Ș* are codul *350*. Astfel numele „Ștefan“ va fi în fruntea listei la sortarea fără opțiuni:

```
sort data-nasterii.txt
Ștefan     : 1984-07-04
Stere      : 1987-10-23
Tomescu    : 1979-03-15
```

Sortarea în ordine inversă nu rezolvă desigur problema:

```
sort -r data-nasterii.txt
Tomescu    : 1979-03-15
Stere      : 1987-10-23
Ștefan     : 1984-07-04
```

Gândiți-vă ce dificultăți generează celelalte litere românești care au diacritice. Mai ales că ele au coduri numerice diferite în diverse sisteme de coduri!⁸

⁸Diferențele acestea de coduri se simt lesne în paginile de web. Renunțarea însă la diacritice în paginile de web nu cred că este o soluție.

3. Extragerea informațiilor

Soluționarea problemei de mai sus depășește cadrul cărții de față. Putem oferi doar o sugestie. Experimentând ca mai sus (sau citind sursa programului *sort*) putem să ne dăm seama că literele alfabetului englez sunt corect sortate. De asemenea, sunt corect sortate cifrele. O idee ar fi să înlocuim pe *S* cu *S1* și pe *Ş* cu *S2*. Acesta ar fi un prim pas. Apoi sortăm și resubstituim pentru a obține literele românești normale.

3.5 Textele ca surse de date

Deși sintagma sună pompos, „baze de date“, în fapt noi am avut de a face parțial în secțiunile anterioare cu baze de date. Pentru a avea de a face cu o bază de date ne trebuie mai întâi un fișier de genul celor în care am înscris cursurile unităților de investiții la un fond mutual. Aceasta este sursa noastră de date. Mai avem nevoie și de un limbaj pe care să-l înțeleagă programe cu care introducem și extragem date din sursa noastră de date. Evident, programe precum *grep* și *sort* joacă doar parțial acest rol. Limbajul programului *grep* este limitat la căutarea cu ajutorul tiparelor. Limbajul pe care-l înțelege *sort* este limitat la operațiuni de sortare.

Un limbaj veritabil de manevrare a surselor de date este *awk*. Creatorii *awk* sunt trei vestiți specialiști în știința computerelor: Alfred Aho, Brian Kernighan și Peter Weinberger. Ei au scris și o carte despre *awk*(v.<<http://cm.bell-labs.com/cm/cs/awkbook/>>).⁹

Pentru a lucra cu *awk* avem mai întâi nevoie de un editor de texte precum Vim, cu care scriem fișierele cu sursele de date. Acestea sunt fișiere de tip text, precum versiunea extinsă a fișierului cu valorile unităților de investiții:

1969-01-31	30
1972-12-31	100
1976-12-31	300
1978-12-31	500
1979-12-31	1000
1981-12-31	700
1984-05-31	3000

Mai avem nevoie, de asemenea, de programul *awk* ca atare. Orice distribuție Linux include acest program. De asemenea, dacă instalați

⁹Mai multe informații despre cei trei puteți găsi pe Wikipedia: despre Aho la <http://en.wikipedia.org/wiki/Alfred_V._Aho>, despre Winberger la <http://en.wikipedia.org/wiki/Peter_J._Weinberger> și despre Kernighan la <http://en.wikipedia.org/wiki/Brian_Kernighan>.

3. Extragerea informațiilor

unelte de programare Unix sub Windows aveți posibilitatea de a folosi *awk*. În exemplele de mai jos *awk* este apelat într-un terminal de tip Unix.

În *awk* avem posibilitatea să folosim limbajul expresiilor regulate. Putem căuta, de pildă, un anumit an în fișierul de mai sus:

```
awk '/1981-[0-9][0-9]-[0-9][0-9]/' curs7.txt
1981-12-31      700
```

Se observă că ziua și luna sunt descrise ca perechi formate din cifre din intervalul care începe cu 0 și se termină cu 9. Am fi putut pune doar anul în expresia regulată, dar lucrul acesta ar fi riscant în cazul unui fișier real, unde anul ar putea coincide cu un șir de cifre aflat pe alt câmp.

Limbajul *awk* ne permite să ne referim la acțiuni. Cea mai simplă acțiune este cea de afișare a conținutului unui câmp. În sursa noastră de date primul câmp conține ani:

```
awk '{print $1}' curs7.txt
1969-01-31
1972-12-31
1976-12-31
1978-12-31
1979-12-31
1981-12-31
1984-05-31
```

Cuvântul-cheie *print* din limbajul *awk* desemnează după cum ne dăm ușor seama o comandă de afișare. Câmpurile sunt indicate prin semnul dolarului urmat de numărul câmpului. În sursa de date câmpurile sunt separate prin tab-uri.

Putem combina tiparele cu acțiunile în felul următor:

```
awk '/1981-[0-9][0-9]-[0-9][0-9]/{print $2}' curs7.txt
700
```

Atenție la semnele de punctuație! Ghilimelele simple închid porțiunea scrisă în limbaj *awk*. Barele oblice cuprind între ele expresia regulată. Acțiunea este indicată de ceea ce este pus între acolade. La început este, evident, invocat *awk*, iar la sfârșit este indicat fișierul cu sursele de date.

Dacă nu se introduce explicit o altă specificație, fiecare rând din sursa de date reprezintă pentru *awk* o înregistrare (*record*, în limba

3. Extragerea informațiilor

engleză). Este vorba desigur despre rândurile logice, separate prin apăsarea tastei pe care scrie „Enter“. Dacă nu se introduce altă specificație, câmpurile unei înregistrări sunt separate prin spații albe (obținute prin apăsarea taste pentru spațiu alb sau a tastei „Tab“).

Putem afla cu ajutorul *awk* numărul de câmpuri din fiecare înregistrare:

```
awk '{print "Numărul de câmpuri: "NF}' curs7.txt
Numărul de câmpuri: 2
Numărul de câmpuri: 2
Numărul de câmpuri: 2
Numărul de câmpuri: 2
Numărul de câmpuri: 2
Numărul de câmpuri: 2
Numărul de câmpuri: 2
```

În comanda *awk* de mai sus ghilimelele duble indică o porțiune de text. Această porțiune este afișată ca atare de către *awk*. Perechea de majuscule *NF* este numele unei variabile interne *awk*. Într-un limbaj de programare o variabilă este precum un sertar sau o cutie în care țin valori. În acest caz, *awk* are din construcție un sertar în care ține valoarea pentru numărul de câmpuri. Această valoare se poate schimba desigur de la o înregistrare la alta. De aici și numele de „variabilă“.

Limbajul *awk* are multe construcții familiare celor care știu alte limbaje de programare. Are însă și multe elemente subtile, bine adaptate lucrului cu sursele de date. Iată, de pildă, cum putem extrage pentru fiecare înregistrare din sursa de date luna din data calendaristică aflată pe primul câmp:

```
awk '{split($1,dc,"-");print dc[2]}' curs7.txt
01
12
12
12
12
12
05
```

După cum am văzut mai sus acțiunile sunt invocate în acolade. Aici avem de a face cu două acțiuni. Prima este aceea de „spargere“ (în engleză *split*) a primului câmp. Ar trebui ceva mai multă familiarizare cu limbajele de programare pentru a înțelege mai exact sintaxa acestei acțiuni. Chiar dacă vă lipsesc asemenea cunoștințe, puteți vedea că

3. Extragerea informațiilor

paranteza rotundă cuprinde trei elemente (numite tehnic „argumente“) ale acțiunii: un prim element indică numărul câmpului, al doilea arată că se va introduce o variabilă botezată „dc“, iar al treilea element indică separatorul folosit în notarea datei calendaristice. Astfel datele calendaristice de pe primul câmp vor fi desfăcute în trei bucăți. Nu rămâne decât să indicăm *awk* a doua acțiune: cea de afișare a lunii (ea fiind a doua „bucată“ din data calendaristică).

Programul *awk* este într-un fel uimitor. El ilustrează felul în care se poate construi o mașinărie complexă de extras date din aparent modeste fișiere de tip text. Mașinăria aceasta este, de pildă, utilizată intens de cei care administrează computere cu sisteme de operare de tip Unix. S-ar putea, de asemenea, ca ea să se ascundă undeva în spatele unor interfețe grafice pline de butoane și casete de dialog.

3.6 Caută și verifică

Procedura folosită atunci când am așezat propozițiile pe rânduri logice poate fi folosită pentru a alcătui lista cuvintelor dintr-un text. Lista respectivă o putem, de pildă, folosi pentru a verifica dacă, într-un alt text, cuvintele care apar acolo sunt scrise ca și cele din listă. Aceasta este operația pe care o facem, în esență, atunci când corectăm un text.

Ceva mai în detaliu, dar tot schițat, planul de acțiune ar arăta astfel. Stabilim modul în care sunt delimitate cuvintele în text. Plasăm apoi fiecare cuvânt pe un rând separat. Sortăm aceste rânduri. Eliminăm repetițiile.

Problema de care ne vom lovi este aceea că o astfel de listă riscă să fie foarte lungă. O putem comprima. Un cuvânt cum este „carte“ are în limba română o serie de forme: „cartea“, „cărți“ etc. Lista este considerabil comprimată dacă fiecare cuvânt din listă este însoțit de un *steag* care ne trimite la un tipar după care pot fi regăsite toate celelalte forme ale cuvântului.

Uneltele pe care le-am discutat deja ne permit construirea unei asemenea liste de cuvinte. Importanța ei nu trebuie subestimată. Poate fi folosită într-un corector ortografic. Poate fi sortată după diverse criterii, pentru a studia structura vocabularului.

Unul dintre cele mai interesante dicționare ale limbii române este *Dicționarul invers* (București: Editura Academiei, 1957). La prima vedere nu pare mare lucru. Pare doar o listă de cuvinte ce se întinde pe mai mult de șapte sute de pagini. La o privire mai atentă șochează bogăția vocabularului limbii române. Ai toate șansele să găsești aproape la tot pasul cuvinte pe care nu le înțelegi. Omul grăbit va strâmba din

3. Extragerea informațiilor

nas pentru că n-are și explicația cuvintelor respective. Utilitatea dicționarului invers este însă alta. Ordonarea cuvintelor din listă este făcută examinând întâi ultima literă. Toate cuvintele care se termină cu litera „a“ vor fi puse la începutul dicționarului. Apoi ele sunt ordonate după a doua literă din finalul lor și așa mai departe. Consultând acest dicționar poți afla care sunt substantivele care se termină în *a* sau care sunt adjectivele care se termină în *e* sau în *re*.

Dicționarul invers este desigur în primul rând un instrument de lucru, o unealtă pentru lingviști. Este însă și un excelent exemplu de management al informației. O simplă ordonare diferită a materialului permite obținerea rapidă de informații. Să culegi toate adjectivele terminate în *e* este o treabă considerabil mai complicată dacă te-ai sluji de un dicționar în care cuvintele sunt ordonate normal, după literele de la începutul lor.

3.6.1 Dicționar pentru corectorul ortografic

Pentru a corecta fișierele text putem, între altele, folosi un corector ortografic creat de Kevin Atkinson și denumit *aspell*¹⁰. Piesa pe care vreau s-o înlocuiesc și să o examinez este chiar lista de cuvinte de care se folosește corectorul.

Aspell este genul de unealtă tipică pentru GNU/Linux. Invocarea sa se face din linia de comandă, precum în cazul de mai jos:

```
aspell --lang ro -c 03-grep.tex
```

La început se tastează numele corectorului. Opțiunea `--lang` este cea care indică limba textului care va fi corectat. Există un mod standard de a nota limbile. În cazul limbii române notația este „ro“. Opțiunea `-c` indică programului că trebuie să verifice (în limba engleză *check*) fișierul denumit „03-grep.tex“. Atenție la faptul că acesta este un fișier de tip text, nu unul binar!

Să vedem acum ce rezultate produce invocarea corectorului. Ceea ce vedeți în figura „Apelarea *aspell*“ este o consolă¹¹ în care apare textul, cu sublinieri, sugestii și un meniu care permite operarea corecturilor.

¹⁰Căutați-l la adresa de Internet <<http://aspell.sourceforge.net/>>. Corectorul *aspell* este parte integrantă din Linux Fedora. Există și versiuni pentru Windows ale corectorului *aspell*, inclusiv o interfață grafică, botezată *freespell* <<http://hcidesign.com/freespell/>>. Recomandabilă este însă utilizarea *aspell* sub GNU/Linux. Precum toate celelalte programe din această distribuție *aspell* este un program cu sursă deschisă (și modificabilă, atâta timp cât respectă licența GNU <<http://www.fsf.org/licensing>>).

¹¹Folosim *urxvt* de la <<http://freshmeat.net/projects/rxvt-unicode/>> pentru a vedea și literele românești.

3. Extragerea informațiilor

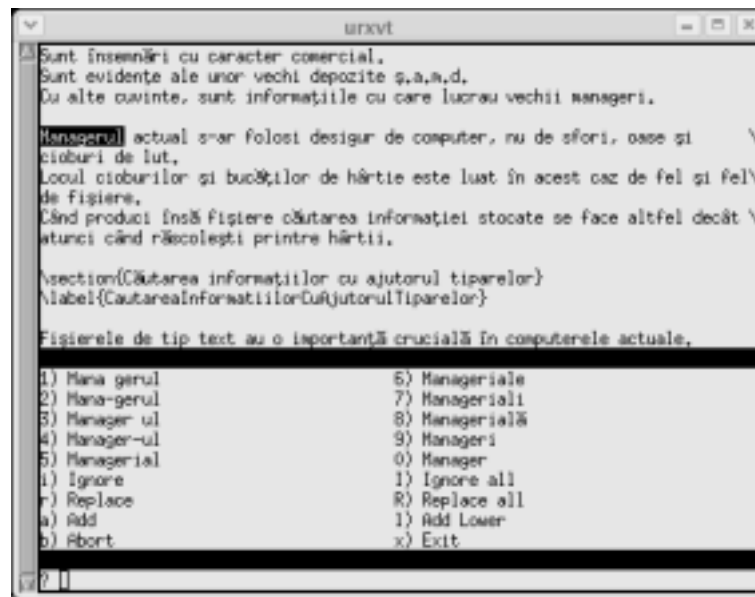


Figura 3.6: Apelarea *aspell*

Aspell execută în fond o dublă căutare: caută fiecare cuvânt din text în dicționarul ortografic; dacă nu găsește cuvântul în dicționar, atunci cercetează dicționarul pentru a găsi forme corecte și cât mai apropiate de cuvântul din text.

Lista necomprimată a dicționarului ortografic cuprinde și forme ale aceluiași cuvânt. În jurul cuvântului „manager“, de pildă, lista poate să cuprindă elemente de genul:

```
management
managementul
managementului
manager
manageri
managerial
managerială
manageriale
managerii
managerilor
```

Forma articulată „managerul“ lipsește din listă. Această absență ridică o problemă interesantă. Putem adăuga forma în „dicționarul personal“. Putem concepe lista comprimată în așa fel încât la expandare să apară și „managerul“. De asemenea, este interesant de văzut dacă unealta pe care o folosim pentru a căuta în listă va sugera o asemenea formă.

3.6.2 Dicționarele bilingve

În forma sa simplă un dicționar bilingv, spre deosebire de dicționarul ortografic, cu prinde două liste puse în corespondență. Unele prezentate în acest capitol permit schimbarea rapidă a listelor între ele. Sortarea după una sau alta dintre liste. De asemenea, o căutare cu *grep* va extrage rândurile pe care se află corespondențele care ne interesează.

De pildă, pe pagina de web a Radei Mihalcea, de la University of North Texas, puteți găsi, sub licență GPL, un dicționar român-englez <<http://lit.csci.unt.edu/~rada/downloads/RoNLP/R.E.tralex>>.

Este interesant să căutăm prin dicționarul Radei Mihalcea cu ajutorul *grep*:

```
grep -hw curent romanian-english.txt
curent <> course
curent <> current
curent <> daily
curent <> draught
curent <> fluent
curent <> fluently
curent <> generally
curent <> instant
curent <> legal
curent <> present
curent <> running
curent <> trend
curent <> usual
curent <> usually
```

Opțiunea *-hw* combină opțiunile *-h*, care cere să nu se afișeze numere de linie, și *-w*, care cere ca tiparul să fie confruntat doar cu cuvinte întregi.¹²

Se observă că Rada Mihalcea folosește drept separator combinația de semne *<>*. Am putea folosi *sort* pentru a sorta după al doilea câmp. Cu ajutorul expresiilor regulate putem identifica două grupuri, despărțite de separator. Este foarte ușor să schimbăm apoi între ele cele două grupuri. Se obține astfel un dicționar englez-român.

Chiar fără a construi dicționarul englez-român putem căuta corespondențele unui cuvânt englezesc cu termenii românești. În căutarea de mai sus ajunge să adăugăm un singur *r* și obținem un efect interesant:

¹²Dacă nu punem opțiunea *-w* ne vom trezi cu toate rândurile în care apare șirul definit în tiparul de căutare.

3. Extragerea informațiilor

```
grep -hw current romanian-english.txt
curent <> current
curs <> current
uzitat <> current
valabil <> current
```

Ideea de a folosi *grep* cu astfel de dicționare am văzut-o prima dată la Frank Richter în programul DING (DIctionary Nice Grep) <<http://www-user.tu-chemnitz.de/~fri/ding/>>. Când îl descărcăm de pe Internet DING are automat instalat un dicționar german-englez. Nu este dificil însă să-i adăugăm dicționarul Radei Mihalcea.

Cu dicționarul român-englez al Radei Mihalcea, căutările cu ajutorul DING arată în felul următor:



Figura 3.7: Căutare în dicționarul român-englez al Radei Mihalcea cu ajutorul DING

Combinând două programe puse sub licența GPL în felul arătat aici, putem obține un instrument de lucru excelent. Licența GPL este foarte importantă în acest caz: ea garantează accesul la surse. Programul obținut este extrem de flexibil. Poate fi adaptat la cele mai diferite întrebări.

Organizația al cărei manager suntem poate folosi deci asemenea unelte simple pentru a rezolva sarcini care par uneori dificile. Se poate folosi forța colectivului și pentru a dezvolta - în situațiile discutate mai sus - dicționare potrivite cu specificul muncii și alte instrumente.

Ideea este însă departe de a fi aceea că exemplele de mai sus au doar o semnificație practică. Ele sunt interesante și pentru felul în care pun în lumină aspectele teoretice ale organizării muncii de stocare și de prelucrare a informației. Dicționarul român-englez al Radei Mihalcea este, de exemplu, un produs colateral al muncii ei de cercetătoare

3. Extragerea informațiilor

în domeniul prelucrării limbajului natural cu ajutorul calculatoarelor. Lucrul acesta se vede, de altfel, din seriile subtile de sinonime pe care le conține dicționarul.

3.7 Extragerea informațiilor: termeni-cheie

- limbajul căutării informațiilor
- expresii regulate
- substituție
- bază de date
- sortare

3.8 Extragerea informațiilor: exerciții și analize de caz

Exercițiul 1: Identificați forme sofisticate de extragere a informațiilor din texte. Folosiți ca punct de plecare datele din Wikipedia, „Information extraction“ <http://en.wikipedia.org/wiki/Information_extraction>.

Analiza de caz 1: Aduceți la o formă standard cazul „Editor modal vs. editor non-modal“. Cercetați varianta cu meniuri a Vim. Când ar fi avantajoase meniurile? Care ar fi pericolele modurilor? Cercetați unul sau mai multe editoare non-modale. Vedeți care sunt avantajele și dezavantajele acestora. Depinde decizia luată de cunoștințe deja existente? De pildă, de faptul dacă stăpânim deja metoda oarbă de dactilografieră. Unde ar fi mai importantă viteza de „dactilografieră“? Nu uitați că metoda oarbă îți permite să te uiți la un text (altul decât cel pe care-l scrii) și să tastezi în același timp. Examinați motivele pentru care decizia în acest caz este una managerială, nu una pur tehnologică.

Exercițiul 2: Descărcați din Biblioteca Gutenberg textul cărții lui Taylor, *The Principles of Scientific Management*. Folosiți versiunea *plain text* sau o altă carte, în funcție de preferințele dumneavoastră, dar tot în format *plain text*. Refaceți exemplele din această unitate de învățare, folosind Vim sau alt editor de texte cu capacități similare.

Exercițiul 3: Dacă nu există deja, instalați *grep* pe computerul dumneavoastră. Vedeți dacă nu găsiți o unealtă similară. Încercați să refaceți, ținând cont de particularitățile uneltei și ale sistemului de operare cu care lucrați, exemplele date aici folosind *grep*.

3. Extragerea informațiilor

Exercițiul 4: Căutați pe Internet unelte de genul celor din Unix, dar pentru Windows. Dacă aveți un sistem Windows și instalați unele dintre aceste unelte veți avea un terminal cu unele capacități similare cu cele ale terminalului din Unix. Un bun punct de plecare Wikipedia, „GnuWin32“ <<http://en.wikipedia.org/wiki/GnuWin32>>. De acolo aveți legături și către proiecte similare cu GnuWin32.

Analiza de caz 2: Particularizați cazul „Limbaaj vs. interfață grafică“ la condiții care vă sunt familiare. Expuneți cazul într-o formă standardizată. Precizați, în condițiile care vă sunt familiare, care ar fi limbajul și care ar fi interfața grafică. Analizați avantajele și dezavantajele fiecărei opțiuni. Fundamentați o decizie benefică din punct de vedere managerial.

Exercițiul 5: Identificați posibilitatea de a folosi limbajul expresiilor regulate în căutările din programele cu care lucrați. Încercați în special să vedeți dacă nu cumva editorul pentru tehnoredactare cu care lucrați poate folosi expresii regulate.

Analiza de caz 3: Discutați cazul „Adrese de e-mail în pagini de web și spam“. Dați-i o formă standard completă. Analizați măsurile luate în firma dumneavoastră pentru a preveni divulgarea către emițători nedorți de mesaje a adreselor de e-mail prin pagina de web. Recurgeți doar la soluții pe care le puteți aplica doar cu forțe proprii? Are sens, din punctul de vedere al raportului costuri-beneficii, să cumprați produsele/serviciile cuiva care să vă rezolve problema? Are vreo importanță înțelegerea de către management a modului în care sunt „capturate“ adrese de e-mail? Există și alte dimensiuni ale problemei decât preluarea din paginile de web a adreselor de e-mail? Construiți eventual cazuri similare.

Exercițiul 6: Căutați pe Internet situri unde se poate învăța limbajul expresiilor regulate. Porniți de la Wikipedia, „Regular Expression“ <http://en.wikipedia.org/wiki/Regular_expression>. Fiți atente/atenți la deosebirile dintre dialectele limbajului.

Exercițiul 7: Încercați să substituiți în Vim cu ajutorul comenzii în linie. Comparați acest mod de a substitui cu înlocuirile pe care le puteți efectua, tot în Vim sau în alt editor, cu ajutorul meniurilor.

Exercițiul 8: [Atenție! Acest exercițiu poate fi foarte periculos. Faceți o copie a fișierului cu care lucrați.] Folosind fișierul cu cartea lui Taylor sau alt fișier (după ce-ați făcut în prealabil o copie) încercați, după modelul din această unitate de învățare, să puneți (aproximativ) fiecare propoziție pe un rând logic. Lucrați întâi cu comenzi fără semnul procentului, pentru a opera doar la nivelul rândului pe care se află cursorul. Studiați rânduri de text diferite. Folosiți *undo* pentru a renunța la rezultatele nemulțumitoare. Studiați cu atenție modul în care

3. Extragerea informațiilor

se determină sfârșitul de rând pe platforma pe care vă aflați. Abia când sunteți siguri că textul comenzii este corect aplicați comanda la nivelul întregului fișier.

Exercițiul 9: Substituția este un mijloc puternic de transformare a unui text, dar periculos. Ordinea substituțiilor este importantă. Studiați pagina de la adresa `<http://www.ub-filosofie.ro/~solcan/wt/gnu/a/agvim.html>`, creată de autorul acestor rânduri. Încercați să înțelegeți rostul ordinii de substituire din scripturile Vim (scripturile Vim sunt fișiere cu comenzi Vim). Dacă, testând practic scripturile, găsiți erori, autorul vă este recunoscător dacă i le comunicați. Încercați să provocați erori deliberat, prin schimbarea ordinii de substituire. Studiați efectul acestor erori.

Exercițiul 10: Încercați să aflați mai multe despre editorul *sed*. Porniți de la Wikipedia, „sed“ `<http://en.wikipedia.org/wiki/Sed>`. Analizați asemănările dintre modul în care se substituie în Vim și modul în care substituim cu *sed*.

Analiza de caz 4: Folosind propria dumneavoastră experiență încercați să construiți cazuri generate de probleme legate de sortare în sensul discutat aici. Nu puneți accent pe problemele pur tehnice. Încercați să identificați ceea ce ține de deciziile de management.

Exercițiul 11: Studiați la adresa `<http://www.ub-filosofie.ro/~solcan/wt/orto/index.html>` pagina creată de autorul acestor rânduri. Cercetați particularitățile instalării dicționarului ortografic pentru limba română disponibil pe pagina menționată.

Exercițiul 12: Citiți articolul lui Rob Reilly, „Extract data from the Internet with Web scraping“ `<http://www.linux.com/articles/52837>`. Analizați uneltele folosite de autor și faceți o comparație cu ceea ce găsiți în această unitate de învățare. Dacă aveți efectiv pe computer programele menționate în articol, încercați să extrageți informații din pagina demonstrativă indicată în articol.

UNITATEA DE ÎNVĂȚARE 4

XML și generalizarea sistemului de marcare a textului

XML este o tehnologie. O tehnologie ale cărei virtuți au fost elogiata în ultimii ani. Din punct de vedere tehnologic, rolul XML este să asigure un transfer sigur și ușor al datelor între programele de computer. XML este o paradigmă a structurării informațiilor în vederea stocării și transferării între aplicații.

Din punct de vedere filosofic se remarcă o serie de intersecții cu zone de mare interes pentru filosofie. Așa este, în primul rând, *semantica*, văzută în filosofie ca un studiu al înțeleșurilor. Se vorbește, de asemenea, în contextul XML, despre *ontologii* - ontologia fiind o veche componentă a filosofiei. Spre deosebire de inteligența computațională, tehnologia XML abordează chestiunile filosofice mult mai prudent, cu un interes exclusiv practic. Aceasta conduce, de altfel, și la o folosire mai degrabă diferită a termenilor amintiți, după cum se va vedea și mai jos.

Din punct de vedere managerial ar fi o mare eroare să ne luăm după elogiile aduse XML. Nu virtuțile tehnologice sunt cele care interesează, ci beneficiile pentru organizație și, până la urmă, pentru noi ca persoane. Din punctul acesta de vedere, XML este o tehnologie cu beneficii în special din perspectiva transparenței și a interoperabilității. Transparența reduce dependența de uneltele de prelucrare ale unei singure firme (cea care ar cunoaște secretele tehnologiei respective). Interoperabilitatea asigură ceva fundamental: fluența circulației informațiilor.

Având în vedere aceste considerente vom prezenta aici doar o mică parte din universul tehnologiei XML. În esență este vorba despre XML ca metalimbaj care permite definirea unei serii întregi de limbaje.

4. XML și generalizarea sistemului de marcare a textului

Obiectivele acestei unități de învățare sunt:

- O prezentare informală a generalizării marcării textelor;
- Explicarea sensului în care se vorbește despre dimensiunea semantică a XML;
- Introducerea unei serii de exemple de aplicații practice ale XML;
- Formularea unei perspective asupra beneficiilor XML pentru programatori, organizații și persoane.

4.1 Generalizarea marcajelor

Pentru a introduce ideea de generalizare a marcajelor vom reveni puțin la fișele noastre de lectură puse în format standard. Să zicem că, într-un fișier de tip text sunt scrise fișe de lectură după schema noastră standard (reținând aici doar elementele obligatorii):

```
Sursa :: ...
Cuvinte-cheie :: ... _conceptul_ ...
Idee-cheie :: ...
```

```
Sursa :: ...
Cuvinte-cheie :: ...
Idee-cheie :: ...
```

```
Sursa :: ...
Cuvinte-cheie :: ...
Idee-cheie :: ...
```

În forma aceasta numai `_conceptul_` are un marcaj explicit. Există însă și un marcaj implicit al textului. Cel puțin pentru noi oamenii este ușor să ne dăm seama unde începe și unde se termină o fișă. Un rând alb urmat de un rând care începe cu `Sursa ::` marchează momentul de început al fișei. La fel, nu este greu să ne dăm seama unde se termină fișa, chiar dacă s-ar adăuga citate, trimiteri, explicații după forma standard.

Forma standard de mai sus este făcută ca să fie citită de oameni. Există doar un mod sistematic de prezentare a informațiilor pentru a ușura manevrarea lor. Dar manevrarea se face tot de către oameni.

Pentru un program de computer (mai corect spus, pentru cine scrie un program de computer) forma de mai sus nu este deloc comodă. De

4. XML și generalizarea sistemului de marcare a textului

exemplu, ar fi mai simplu să se știe explicit unde începe și unde se termină fișa, unde începe și unde se termină sursa ș.a.m.d. Lucrurile ar putea arăta precum în schema următoare:

```
<fișa>
  <sursa> ... </sursa>
  <cuvinte-cheie> ... </cuvinte-cheie>
  <idee-cheie> ... </idee-cheie>
</fișa>
<fișa>
  <sursa> ... </sursa>
  <cuvinte-cheie> ... </cuvinte-cheie>
  <idee-cheie> ... </idee-cheie>
</fișa>
<fișa>
  <sursa> ... </sursa>
  <cuvinte-cheie> ... </cuvinte-cheie>
  <idee-cheie> ... </idee-cheie>
</fișa>
```

Toată această unitate de învățare se învârtă în jurul avantajelor și dezavantajelor unei astfel de scheme de structurare a informațiilor.

Un lucru este cert: generalizarea marcării textului (în stil XML mai sus) este un mijloc de a stoca *explicit* mult mai multe informații. Problema care este rezolvată, cum spuneam mai sus, este aceea de a oferi informațiile într-o formă comodă pentru prelucrarea lor de către programele de computer. În literatura despre XML veți citi despre dimensiunea *semantică* a acestor marcaje. Dar când lipsesc marcajele? Dispare semantica? Nu - cel puțin pentru oameni. Pentru a desemna situația respectivă am folosit *ad hoc* termenul de „criptosemantică“. Cazul în fața căruia ne-am afla, în forma sa generală, ar fi:

__Criptosemantică vs. semantică__

Oamenii sesizează cu ușurință înțelesurile unui text a cărui limbă o cunosc.
Din perspectiva unui computer,
„a cunoaște o limbă naturală“ înseamnă a rezolva o dificilă problemă de inteligență computațională.

==Problema==

4. XML și generalizarea sistemului de marcare a textului

Prelucrarea unui text cu ajutorul computerului presupune ușurința extragerii informațiilor.

Un text într-o limbă naturală are însă o „criptosemantică“ pentru un program de computer.

Chiar dacă textul cuprinde date numerice sau alt gen de informații care nu țin de limba naturală, programul de computer întâmpină probleme dacă informația semantică nu este structurată astfel încât să fie lesne extrasă de către program.

[XML rezolvă problema cu ajutorul unui sistem de marcaje generalizate. Aceste marcaje sunt *extensibile* (de aici „X“-ul din numele XML).]

4.2 Exemple de aplicații care utilizează XML

Dacă ne interesează nu tehnologia ca atare, ci *ceea ce poate ea face pentru noi*, exemplele de utilizare sunt mai folositoare decât analiza trăsăturilor tehnologiei respective. Vom oferi de aceea câteva exemple - dintre cele familiare autorului - de utilizare a XML.

4.2.1 Situl Perseus

Situl Perseus¹ a fost creat pentru „a aduce studiile clasice în lumea digitală“. Pe acest sit veți găsi în special texte în limba greacă veche și în limba latină.

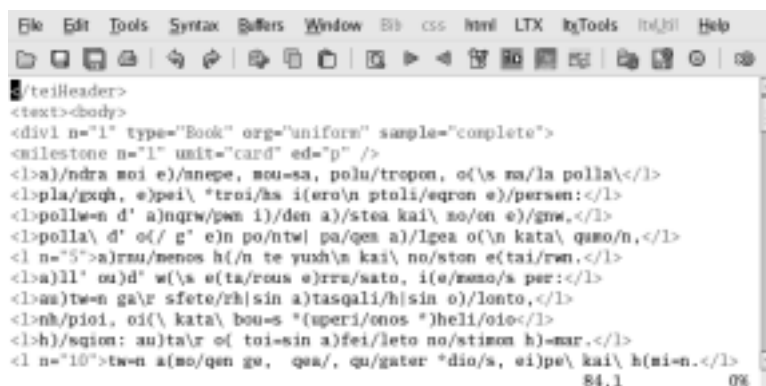
În versiunea recentă a sitului Perseus sursele textelor pe care le vedeți afișate pe ecran sunt într-un format care folosește XML. Fișierul XML al unora dintre aceste texte poate fi descărcat de pe situl Perseus. În figura „Fragment din Odiseea“ se vede chiar începutul textului homeric.

Marcajele folosite în textul homeric nu sunt foarte complicate. În bună parte sunt extrem de transparente. Este limpede, de pildă, că în figura „Fragment din Odiseea“ începe corpul textului. Versurile sunt prinse într-o pereche de marcaje `<1>` `</1>` (unde „1“ vine de la englezescul *line*). Ca și-n edițiile obișnuite, versurile sunt numărate din cinci în cinci.

Mai ciudat poate să pară modul în care este scris textul grecesc. Asta n-are nici o legătură cu XML. În XML se poate folosi *unicode*, ceea ce nu înseamnă altceva decât că ar fi putut fi acolo literele grecești. Sau am putea pune noi în text litere grecești.

¹<http://www.perseus.tufts.edu/hopper/>

4. XML și generalizarea sistemului de marcare a textului



```
File Edit Tools Syntax Buffers Window BB CSS HTML LTX TeXTools TeXUI Help
</teiHeader>
<text><body>
<div1 n="1" type="Book" org="uniform" sample="complete">
<milestone n="1" unit="card" ed="p" />
<l>a)/ndra moi e)/annepe, nou-sa, polu/trepon, o(\s na/la polla</l>
<l>pla/gxqh, e)pei\ *troi/hs i(ero\n ptoli/sqron e))/persen:</l>
<l>polla-n d' a)ngwr/pwn i)/den a)/stea kai\ no/en e)/gwr.</l>
<l>polla\ d' o(/ g' e)n po/ntw\ pa/qen a)/lgea o(\n kata\ qamo/n,</l>
<l n="5">a)rru/menos h(/n te yush\h kai\ no/ston e(tai/rwn.</l>
<l>a)ll' ou)d' w(\s e(ta/rous e)rru/sato, i(e/meno/s per:</l>
<l>a)twen ga'r sfete/zhsin a)tasqali/hjsin o)/lonto.</l>
<l>nh/pioi, oi(\ kata\ bou-s *(uperi/enos *)heli/oi</l>
<l>h)/sqion: au)ta'r o(\ toi-sin a)fei/leto no/stimon h)-mar.</l>
<l n="10">twen a(mo/qen ga, qea/, qu/gater *dio/s, ei)pe\ kai\ h(mi-n.</l>
84,1 0%
```

Figura 4.1: <http://www.perseus.tufts.edu/hopper/>

Opțiunea pentru sistemul de transcriere cu litere latine a textului grecesc are avantajele sale. De exemplu, este mai lesne să scrii, bunăoară în editorul Vim, o expresie cu ajutorul căreia să cauți un cuvânt sau un fragment de text care are un anumit tipar. În figura „Căutare în textul Odiseei” apare o căutare după începutul cuvântului grecesc pentru „sirenă”. Episodul găsit este cel în care Circe îl previne pe Ulise asupra pericolului reprezentat de sirene.



```
File Edit Tools Syntax Buffers Window BB CSS HTML LTX TeXTools TeXUI Help
<l>*seirh-nas no/n prw-ton a)fi/ceai, ai(/ r(a/ te pa/ntas</l>
<l n="40">a)ngwr/pous qe/lgousin, o(/tis sfeas ei)safi/khtai.</l>
<l>(s tis a)idrei/h\ pela/sh\ kai\ fqo/ggon a)kou/sh</l>
<l>*seirh/nen, tw=| d' ou)/ ti gunh\ kai\ nh/pia te/kna</l>
<l>oi)/kade nosth/santi pari/statai os)de\ ga/nuntai.</l>
<l>alla/ te *seirh-nes ligurh=| qe/lgousin a)oidh=|</l>
<l n="45">h(/menai e)n leime-ni, pols\ s d' a)mf' o)steo/fin qi\s</l>
<l>a)ndrw-n pugone/nen, peri\ de\ r(inoi\ minu/qousi.</l>
<l>alla\ parecela/an, e)pi\ d' ou)/at' a)lei-yai e(tai/rwn</l>
<l>khro\n deyh\sas melikde/s, mh/ tis a)kou/sh</l>
<l>twen a)/llon: a)ta'r au)to\s a)koue/men ai)/ k' e)qe/lh\sqa.</l>
<l n="50">dhsa/ntwn s' e)n nhi\ qoh=| xei-ra/s te po/das te</l>
<l>e)rqo/s e)n i(stope/dh\, e)k d' au)ton= pei/rat' a)nh/fqw.</l>
<l>e)/fra ke terpo/menos o)/p' a)kou/sh)s *seirh/noiin.</l>
[BISFER-1][FORMAT=dos] [TYPE=XML] [POS=6043,1]
/seirh
```

Figura 4.2: Căutare în textul Odiseei

O dată identificat un punct din text este desigur ușor de stabilit și unde ne găsim în textul homeric. În general, fișierul XML poate cuprinde o sumedenie de informații despre text. Nu toate apar, ca să spunem așa, la suprafață, atunci când fișierul este vizualizat cu ajutorul unui program care-i dă o formă estetică, plăcută la lectura normală.

4. XML și generalizarea sistemului de marcare a textului

4.2.2 DocBook

XML, ca metalimbaj, permite specificarea altor limbaje. Un asemenea limbaj este DocBook, care permite crearea de documente a căror sursă este în format XML.

Cu ajutorul DocBook este creată o sursă în care sunt puse toate informațiile cu privire la *conținutul* documentului. Sunt posibile apoi diverse transformări care duc la realizarea unui document *tehnoredactat* în funcție de destinația (tipărire, pagină de web etc.).

Cu DocBook pot fi, între altele, create cărți. Editura O'Reilly a fost prima care a utilizat acest format pentru cărțile sale.

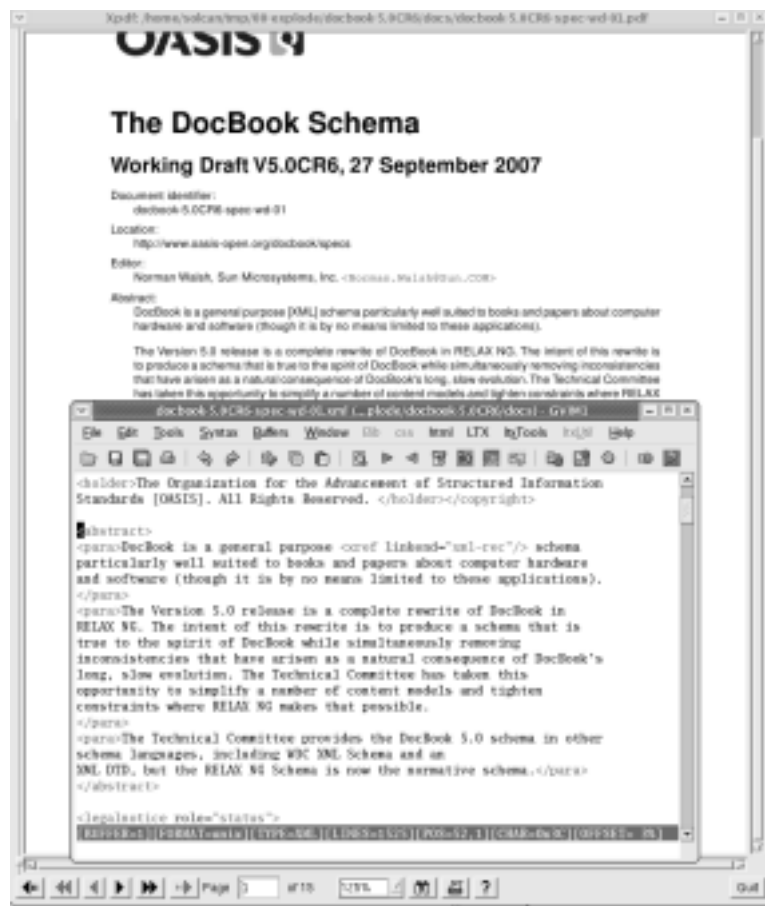


Figura 4.3: Document scris în DocBook

În figura „Document scris în DocBook“ se vede în prim plan sursa.

4. XML și generalizarea sistemului de marcare a textului

În plan secund este documentul pregătit pentru tipărire sau lectura pe ecran. Fragmentul din sursă cuprind chiar rezumatul documentului.²

4.2.3 Formatul OpenDocument

Dacă instalați OpenOffice2 și scrieți un text, formatul nativ al acestuia este odt. Ca atare, odt este o arhivă. O puteți deschide cu dezarhivatorul cu care deschideți fișierele zip.

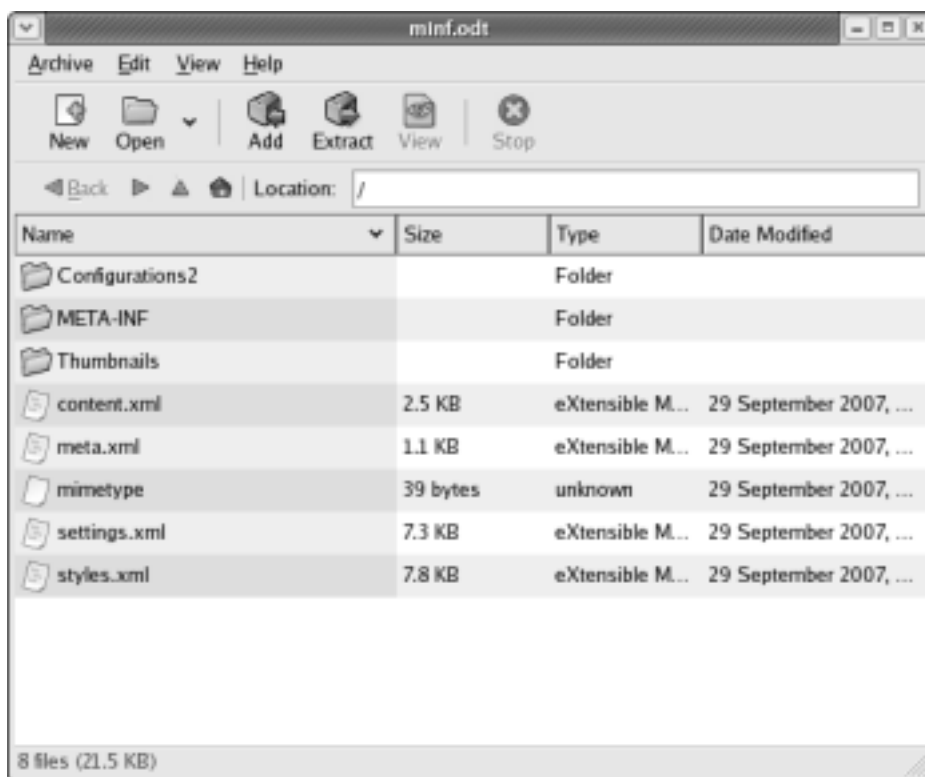


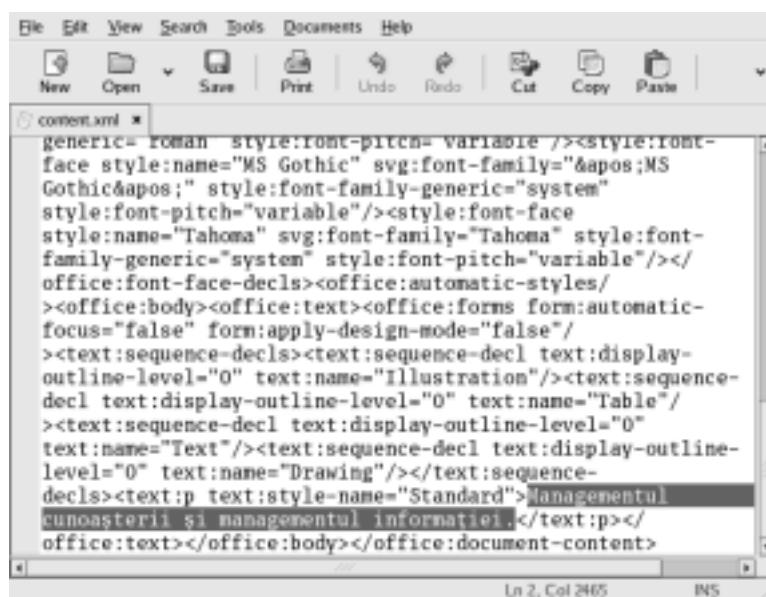
Figura 4.4: Un fișier odt dezarhivat

Tehnologia folosită în cazul odt este XML. Putem deschide și cu un editor de texte fișierul unde se află conținutul textului.

În figură se vede că OpenOffice2 a pus foarte multe informații în fișier. Ceea ce a fost scris de la tastatură este doar textul scos în evidență prin sublinierea de către noi a fondului.

²Pentru alte informații și pentru documentul „DocBook V5.0CR6“ ca atare a se vedea <<http://www.docbook.org/>>.

4. XML și generalizarea sistemului de marcare a textului



```
File Edit View Search Tools Documents Help
New Open Save Print Undo Redo Cut Copy Paste
content.xml
generic=roman style:font-pitch=variable /><style:font-
face style:name="MS Gothic" svg:font-family="&apos;MS
Gothic&apos;" style:font-family-generic="system"
style:font-pitch="variable"/><style:font-face
style:name="Tahoma" svg:font-family="Tahoma" style:font-
family-generic="system" style:font-pitch="variable"/></
office:font-face-decls><office:automatic-styles/
><office:body><office:text><office:forms form:automatic-
focus="false" form:apply-design-mode="false"/
><text:sequence-decls><text:sequence-decl text:display-
outline-level="0" text:name="Illustration"/><text:sequence-
decl text:display-outline-level="0" text:name="Table"/
><text:sequence-decl text:display-outline-level="0"
text:name="Text"/><text:sequence-decl text:display-outline-
level="0" text:name="Drawing"/></text:sequence-
decls><text:p text:style-name="Standard">Managementul
cunoașterii și managementul informației.</text:p></
office:text></office:body></office:document-content>
```

Figura 4.5: Fișier cu conținutul textului

4.2.4 RSS

Multe situri ale agențiilor de presă, ziarelor oferă „știri pe bandă“. Cu ajutorul unui program special ne putem conecta la sursa de știri. Nu trebuie decât să găsim pe pagina de web butonul RSS.

Un asemenea buton RSS există și pe pagina de web a autorului acestei unități de învățare. Rostul său nu este acela de a furniza „știri“ referitoare la schimbările intervenite în pagina de web sau pur și simplu de a indica legături interesante pe Internet.

Sursa RSS este scrisă în format XML. Mai nou, asemenea pagini XML pot fi văzute și navigatorul de Internet Firefox. În figura „RSS în Firefox“ se vede sursa RSS și felul în care apare textul în pagina de web a autorului.

4.3 Avantajele și dezavantajele XML

XML oferă programatorilor un mod de a reprezenta datele structurat. Avantajele sale esențiale țin de posibilitatea de a prelucra mai ușor informația și de a o transfera mai lesne de la o aplicație la alta.

XML este scris în fișiere de tip text. Acestea pot fi citite cu orice editor de texte. Prin urmare, structura acestor fișiere este transparentă, accesibilă oricui. Fișierele XML nu arată de parcă ar fi criptate.

4. XML și generalizarea sistemului de marcare a textului



Figura 4.6: RSS în Firefox

Mai nou, Word2007 are drept format nativ al fișierelor sale tot XML. S-ar putea ca pe viitor să dispară situația de pe vremea fișierelor doc binare al căror format nu era cunoscut decât de programatorii firmei sau de colaboratorii acesteia. Orice altă firmă va putea scrie programe care deschid fișiere Word2007 și scriu fișiere Word2007.

Examinând figurile din această unitate de învățare se văd și dezavantajele XML. Marcajele XML sunt logoreice. Se vorbește mult în XML. Din această pricină, chiar dacă sunt de tip text, fișierele XML nu sunt ușor de citit sau de scris cu un editor de texte. Este nevoie de editoare speciale pentru XML.

XML nu este genul de limbaj în care să le fie ușor oamenilor să-și scrie textele. Prin contrast, LaTeX este un limbaj în care se poate scrie direct de către oameni, ca-n orice limbaj de programare.

4.4 XML și generalizarea sistemului de marcare a textului: termeni-cheie

- XML
- Criptosemantică vs. semantică
- DocBook

4. XML și generalizarea sistemului de marcare a textului

- OpenDocument
- RSS

4.5 XML și generalizarea sistemului de marcare a textului: exerciții și analize de caz

Exercițiul 1: Clarificați conceptul de limbaj de marcare. Folosiți drept punct de plecare articolul „Markup language“ din Wikipedia <http://en.wikipedia.org/wiki/Markup_language>.

Exercițiul 2: Stabiliți principalele legături dintre evoluția XML și dezvoltarea Internetului. Folosiți drept punct de plecare articolul despre XML din Wikipedia <<http://en.wikipedia.org/wiki/XML>>.

Exercițiul 3: Folosind ca punct plecare Wikipedia, „Semantic Web“ <http://en.wikipedia.org/wiki/Semantic_Web> clarificați noțiunea de *web* semantic. Ce rol joacă XML în această transformare a sistemului de pagini de pe Internet?

Analiza de caz 1: Folosind informațiile culese în exercițiile anterioare dați o formă standard completă cazului „Criptosemantică vs. semantică explicită“. Analizați în special avantajele și dezavantajele semanticii explicite? Va deschide, de pildă, aceasta mai multe posibilități pentru cenzură?

Analiza de caz 2: Căutați un caz de utilizare practică a XML. Analizați, după schema standard, cazul respectiv.

UNITATEA DE ÎNVĂȚARE 5

SQL din perspectiva fluxului informațiilor

Dacă citim o carte de informatică pentru managementul firmei, precum cea scrisă de profesorul Ioan Radu și colaboratorii săi[8], găsim acolo o descriere conceptuală a fluxurilor care asigură aducerea la zi a informațiilor și obținerea de informații din baza de date. Nici un program din cele descrise în cartea amintită nu se descurcă de fapt fără a avea acces la informații. O bună bază de date este esențială pentru funcționarea oricărei organizații moderne.

Felul în care tehnologia care stă în spatele SQL, exemplul pe care se va axa această unitate de învățare, „macină“ informațiile este mai puțin semnificativ aici. Din perspectiva managementului informației, tehnologia SQL este semnificativă ca paradigmă a unei verigi extrem de importante pe fluxul informațiilor. Operăm cu o sumedenie de informații pe care dacă le-am stoca doar în fișiere de genul celor de tip *doc* sau *pdf* am fi nevoiți să le tot copiem dintr-un document într-altul sau, mai rău, să tot retastăm în computer textul informațiilor respective.

De asemenea, aici este vorba și despre managementul fluxului de informații la nivel de întreprindere, la nivelul deci al unei organizații mari, care operează cu foarte multe informații. Și în sensul acesta SQL este paradigmatic.

Conținutul unității de învățare nu presupune însă că avem acces la o mare bază de date și la programul folosit pentru a lucra cu ea. Lumea marilor baze de date este redusă, din motive de ordin didactic, la dimensiuni „casnice“. Tehnicile explicate, acolo unde are sens, ar putea fi folosite însă și-n practică.

Obiectivele acestei unități de învățare sunt:

- Familiarizarea cu unele dintre elementele de bază ale SQL;

- Înțelegerea modului în care informația este stocată și regăsită cu ajutorul SQL;
- Cunoașterea modului în care accesul la baza de date poate fi automatizat în vederea aducerii la zi, pe întregul flux, a informației.

5.1 SQL pe computerul de acasă

Studiind capitolele anterioare v-ați dat desigur seama că pentru a interacționa cu o sursă de date trebuie să folosim un limbaj. Pentru bazele de date „de întreprindere“ există un limbaj standard: SQL. Numele limbajului vine de la inițialele componentelor sintagmei „*Structured Query Language*“. Adevărul este că SQL e mai mult decât un limbaj de căutare în sursele de date. Toate bazele de date importante înțeleg SQL și poți face o sumedenie de operații cu ajutorul SQL.

Problema cea mare a bazelor de date „de întreprindere“ este că sunt destul de greu de instalat acasă sau pe computerul de la serviciu pe care am vrea să învățăm SQL. În principiu, lucrul acesta este posibil - chiar și pe un sistem Windows98. Administrarea bazei de date este însă dificilă (mai ales în cazul începătorilor sau al persoanelor care vor să-și facă doar o idee despre SQL). Soluția pe care o propunem aici este folosirea SQLite.

Autorul principal al SQLite este D. Richard Hipp. Marele avantaj al SQLite este că sursa de date este într-un singur fișier și nu trebuie să pornești nici un fel de server pentru a avea acces la ea. Este o soluție excelentă pentru cei care învață SQL. Este, de asemenea, o soluție de luat în seamă în practică: pentru bazele de date ale paginilor de web, de exemplu. În întreprindere trebuie să fim însă atenți la problemele de securitate a datelor: cine are acces la fișierul cu datele stocate de SQLite are acces la toate datele. Accesul nu este selectiv.

Pentru învățarea SQL (dacă porniți de la zero) cartea scrisă de Ben Forta[3] este excelentă. Nu folosiți însă exemplele din traducerea românească. Cel puțin versiunea din 2002, cea pe care am citit-o, este plină de erori. Trebuie să știi deja SQL ca să folosești exemplele de acolo. Descărcați exemplele de pe situl autorului <<http://www.forta.com/books/0672325675/>>. În limita capacităților SQLite, exemplele acestea merg fără probleme; dacă nu știți decât prea puțină engleză, traducerea românească este totuși utilă pentru a prinde sensul exemplelor.

Programul SQLite se poate descărca de pe Internet. Exemplele date în continuare sunt lucrate cu sqlite-2.8.16. Această versiune a SQLite este în principal și-n atenția lui Chris Newman în cartea sa despre

5. SQL din perspectiva fluxului informațiilor

SQLite[7].¹ SQLite este disponibilă atât pentru Linux, cât și pentru Windows.

Instalarea programului SQLite nu este deloc complicată. Trebuie să fiți doar atente și atenți să aveți acces la executabilul `sqlite`. În Unix aceasta înseamnă că puteți purta un dialog de acest gen:

```
which sqlite
/usr/bin/sqlite
```

Citiți însă cu grijă instrucțiunile de instalare.

Programul SQLite ca atare este un instrument apelat prin intermediul comenzii în linie (într-o consolă Unix sau o fereastră MSDos).

5.2 Un exemplu de bază de date SQLite

Conform principiului acestui manual ar trebui ca fiecare să lucreze cu exemple familiare. Cum autorul lucrează într-o universitate, unele dintre „articolele de inventar“ cel mai frecvent întâlnite sunt cărțile. Așa că exemplul dat în continuare se referă la cărți. Puteți vedea însă alte exemple în cartea lui Forta[3] sau în alte cărți despre SQL.

Cum spuneam, în cazul SQLite baza de date stă într-un singur fișier. Putem deschide cu Vim sau alt editor de texte acest fișier, dar vom găsi un amestec de text lizibil și de semne ininteligibile. Fișierul respectiv nu este un fișier de tip text. O idee despre ceea ce se vede în Vim în cazul acestui gen de fișiere ne putem face cu ajutorul figurii „Fișierele SQLite nu sunt de tip text“.

Crearea și introducerea de date în aceste fișiere, după cum vom vedea mai jos, se face însă cu ajutorul unor fișiere de tip text. Am ales drept nume pentru fișierul în care SQLite stochează datele „bib.db“ pentru că „bib“ amintește de cărți, iar „db“ este extensia standard pentru aceste baze de date.

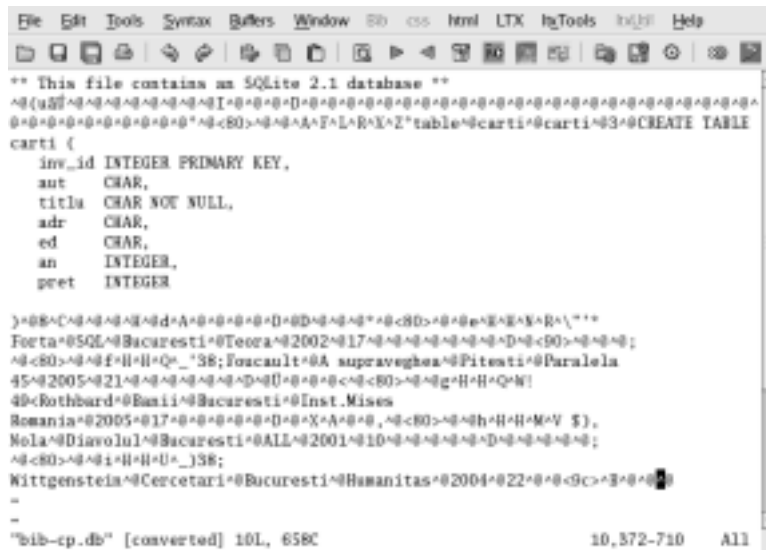
Ne slujim de un fișier *bib.sql*, precum cel din figura „Crearea unei baze de date SQLite“, pentru a genera baza de date și structurile în care vor fi inserate efectiv datele.

Pentru a crea fișierul ca atare și structurile sale folosit comanda în linie (comanda dată în consolă sau fereastră MSDos):

```
sqlite bib.db < bib.sql
```

¹Există și `sqlite3`. Versiunea 2 are numai un caracter „istoric“, dar este încă utilă - cel puțin pentru scopuri didactice.

5. SQL din perspectiva fluxului informațiilor



```
File Edit Tools Syntax Buffers Window DB CSS HTML LTX InTools InDB Help
** This file contains an SQLite 2.1 database **
CREATE TABLE carti (
  inv_id INTEGER PRIMARY KEY,
  aut CHAR,
  titlu CHAR NOT NULL,
  adr CHAR,
  ed CHAR,
  an INTEGER,
  pret INTEGER
);
INSERT INTO carti (aut, titlu, adr, ed, an, pret) VALUES ('Forta', 'Teora', '2002', '17', 'Teora', '17');
INSERT INTO carti (aut, titlu, adr, ed, an, pret) VALUES ('Foucault', 'supraveghea', 'Pitești', 'Paralela 45', '2005', '21');
INSERT INTO carti (aut, titlu, adr, ed, an, pret) VALUES ('Rothbard', 'Banii', 'București', 'Inst. Mises Romania', '2005', '17');
INSERT INTO carti (aut, titlu, adr, ed, an, pret) VALUES ('Nola', 'Diavolul', 'București', 'ALL', '2001', '10');
INSERT INTO carti (aut, titlu, adr, ed, an, pret) VALUES ('Wittgenstein', 'Cercetari', 'București', 'Humanitas', '2004', '22');
--
--
"bib-cp.db" [converted] 10L, 658C 10,372-710 All
```

Figura 5.1: Fișierele SQLite nu sunt de tip text

Pentru a pregăti date în vederea introducerii în tabelul *carti* construim un alt fișier *sql*, botezat *bibdata.sql*. În figura „Fișier folosit la introducerea datelor“ se poate vedea cum arată datele astfel pregătite.

Din nou, introducerea datelor ca atare se face printr-o comandă în linie:

```
sqlite bib.db < bibdata.sql
```

Modul de generare a bazei de date și de introducere a informațiilor descris mai sus este rapid și sigur. Există însă și posibilitatea, după cum se vede în figura „Tabel cu cărți“ de a vizualiza conținutul bazei de date și în mod grafic, nu doar în mod text.

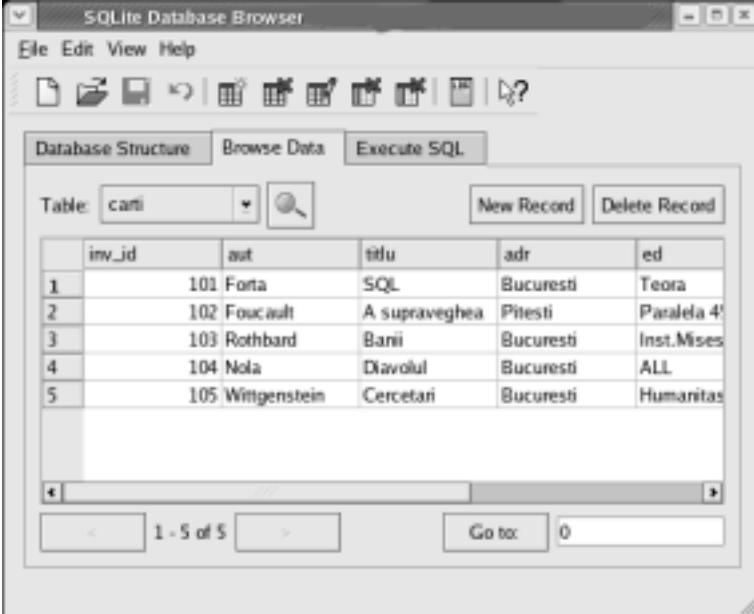
```
CREATE TABLE carti (  
    inv_id INTEGER PRIMARY KEY,  
    aut    CHAR,  
    titlu  CHAR NOT NULL,  
    adr    CHAR,  
    ed     CHAR,  
    an     INTEGER,  
    pret   INTEGER  
);
```

Figura 5.2: Fișier pentru crearea unei baze de date SQLite

```
INSERT INTO carti (inv_id, aut, titlu, adr, ed, an, pret)  
VALUES {101, 'Forta', 'SQL', 'Bucuresti', 'Teora', 2002, 17};  
  
INSERT INTO carti (inv_id, aut, titlu, adr, ed, an, pret)  
VALUES {102, 'Foucault', 'A supraveghea', 'Pitesti', 'Paralela 45', 2005, 21};  
  
INSERT INTO carti (inv_id, aut, titlu, adr, ed, an, pret)  
VALUES {103, 'Rothbard', 'Bani', 'Bucuresti', 'Inst.Mises Romania', 2005, 17};  
  
INSERT INTO carti (inv_id, aut, titlu, adr, ed, an, pret)  
VALUES {104, 'Nola', 'Diavolul', 'Bucuresti', 'ALL', 2001, 10};  
  
INSERT INTO carti (inv_id, aut, titlu, adr, ed, an, pret)  
VALUES {105, 'Wittgenstein', 'Cercetari', 'Bucuresti', 'Humanitas', 2004, 22};  
"
```

Figura 5.3: Fișier folosit la introducerea datelor

5. SQL din perspectiva fluxului informațiilor



The screenshot shows the SQLite Database Browser interface. The 'Browse Data' tab is active, displaying a table named 'carti'. The table has five columns: 'inv_id', 'aut', 'titlu', 'adr', and 'ed'. There are five rows of data. The interface includes a menu bar (File, Edit, View, Help), a toolbar with various icons, and buttons for 'New Record' and 'Delete Record'. A status bar at the bottom shows '1 - 5 of 5' and a 'Go to:' field with the value '0'.

	inv_id	aut	titlu	adr	ed	
1		101	Forta	SQL	Bucuresti	Teora
2		102	Foucault	A supraveghea	Pitesti	Paralela 4'
3		103	Rothbard	Banii	Bucuresti	Inst.Mises
4		104	Nola	Diavolul	Bucuresti	ALL
5		105	Witgenstein	Cercetari	Bucuresti	Humanitas

Figura 5.4: Tabel cu cărți

5.3 Limbajele de calculator și fluența circulației informațiilor

Când eram copil, în timpul unei excursii cu școala, am fost să vizităm o fabrică de rulmenți. Era construită de către specialiști veniți din Japonia. S-ar putea ca, după atâta timp, memoria să mă înșele. Nu-mi amintesc însă decât de un singur moment în care piesa prelucrată era întoarsă cu mâna. În rest totul făceau mașinile. Erau încă și muncitori japonezi în fabrică; nu făceau însă altceva decât să urmărească imposibil procesul tehnologic. Pe întregul flux tehnologic nu era decât o singură operație executată manual. Restul era automatizat.

Fluxul informațiilor nu este vizibil precum acela al fabricării rulmenților cu bile. Dar importanța fluenței circulației informației s-ar putea să fie chiar mai mare. Operațiile efectuate manual întârzie considerabil procesul și introduc erori. Pentru a corecta erorile este iar nevoie de o intervenție manuală și acest gen de management al informației este extrem de deficitar.

Acolo unde intervențiile manuale decurg după un tipar ele trebuie înlocuite cu procese programate. Limbajele de programare folosite pot fi foarte variate. SQL este aici doar un simplu exemplu. Este un limbaj important, dar specializat pentru baze de date. E semnificativ însă pentru că ne întâlnim adesea cu el. Poate nu ne dăm seama, dar SQL este o parte a mașinării unor pagini de web pe care s-ar putea să le vizităm adesea. Vom explica și noi mai jos felul în care se produce această integrare.

Nu există la drept vorbind un singur limbaj cu care să asiguri fluența circulației informației. Există, de asemenea, și alegeri ce țin de gust. Dacă este vorba de un limbaj de uz general pentru acest tip de sarcini, autorul are o preferință pentru Python.²

Să ne gândim, de exemplu, că datele despre cărțile din baza de date SQLite ar fi existat deja într-un fișier XML. Dacă o operație de trecere de la fișierul respectiv la unul precum cel din figura „Fișier folosit la introducerea datelor“ are loc în mod repetat, atunci are sens să fie scris un program care să facă el acest lucru. Programul respectiv ar putea fi scris în Python sau în alt limbaj, în funcție de condiții.

Din perspectivă strict managerială, pentru o firmă are sens să angajeze cel puțin o programatoare sau un programator care să asigure

²Pentru Python puteți consulta lecțiile dedicate limbajului Python de Uta Priss, lecții concepute ca un curs de „Computer Programming for Information Management“ <<http://www.upriss.org.uk/python/PythonCourse.html>>. Cursul a fost ținut la School of Library and Information Science, Indiana University Bloomington.

5. SQL din perspectiva fluxului informațiilor

tocmai această fluidizare a circulației informației. Este o mare iluzie să crezi că sunt suficiente marile complexe de programe, veritabile uzine de prelucrat informații. Totdeauna mai este ceva de „cârpit“ și doar o persoană care are o cunoaștere personală a situației poate face asta.

Ideea anterioară este valabilă chiar dacă este vorba despre programe obișnuite de birotică. Este greu să i se ceară personalului nespecializat să scrie programele de care este nevoie.³

Și ideea inversă este adevărată. N-are sens să cumpărați pentru firmă o pagină de *web* strălucitoare, dar nepotrivită cu forțele de care dispuneți în firmă. Treptat vor fi multe de „meșterit“ la pagina de *web*. Ea se va învechi oricum, iar în unele puncte circulația informației se va deregla și se va osifica. Așa că are sens s-o adaptați la resursele (financiare și umane) de care dispuneți.⁴

5.4 Căutarea în baza de date

Este important să ne dăm seama că fișierul *bib.db* este o jucărie. Are un singur tabel și cinci înregistrări. O bază de date reală poate să fie imensă. În timp ce, dacă ne uităm în figura „Tabel cu cărți“, ne facem rapid o idee despre datele din tabel, într-un tabel real acest lucru n-ar fi posibil. Aici ne vine în ajutor limbajul SQL.

Instrucțiunile SQL sunt destul de transparente pentru cine știe limba engleză. În continuare deschidem sursa de date *bib.db* cu ajutorul *sqlite* și cerem să fie afișat conținutul câmpului *aut*:

```
$ sqlite bib.db
SQLite version 2.8.16
Enter ".help" for instructions
sqlite> SELECT aut
...> FROM carti;
```

³Microsoft Office, care este prezent în multe birouri, inclusiv în varianta cu bază de date (și limbaj SQL) este deseori exploatat necorespunzător. Asta pentru că și lucruri care par simple, precum arhitectura datelor, nu sunt lesne de făcut fără o diviziune a muncii. Iar pentru limbajul din Microsoft Office ce poate fi folosit pentru îmbunătățirea lucrului cu această aplicație se poate consulta, de către începători sau un public mai larg, Paul McFedries, *VBA - Ghid pentru începători* (București: Teora, 2006).

⁴Nu pot, de pildă, să încerc să fac să semene pagina mea personală de *web* cu aceea de la UEFA. Și aceea se mai dereglează (o vreme cel puțin, pe pagina lor în japoneză, caracterele ba apăreau, ba dispăreau după câte o reîmprospătare automată a paginii). N-are sens nici să concurez cu situl Perseus. Pagina UEFA are o grafică flamboaiantă, iar situl Perseus o bogăție incredibilă a informațiilor academice furnizate. Numai „meșteritul“ (reparatul pe ici pe colo, îndreptarea unor disfuncționalități etc.) mi-ar depăși rapid cunoștințele și timpul de care dispun.

5. SQL din perspectiva fluxului informațiilor

Forta
Foucault
Rothbard
Nola
Wittgenstein

Observați că apare un prompt *sqlite>* specific programului de administrare a bazei de date. După *SELECT aut* am apăsat tasta „Enter“. De asemenea, după *FROM carti*; am apăsat tasta „Enter“. Nu uitați să puneți punct și virgulă!

Programul *sqlite* are o istorie a comenzilor; dacă apăsați pe tasta cu săgeata în sus sau pe cea cu săgeata în jos puteți naviga printre comenzi. N-are rost să tastați de mai multe ori aceeași comandă. De asemenea, puteți transforma lesne o comandă într-altă comandă. Istoria comenzilor funcționează și după promptul *...>* ceea ce este o idee extraordinar de bună.

Putem selecta și mai multe câmpuri o dată. De exemplu, dacă vrem să vedem așa-numitele fapte legate de publicare (locul, editura, anul) procedez în felul următor:

```
sqlite> SELECT adr, ed, an
...> FROM carti;
Bucuresti|Teora|2002
Pitesti|Paralela 45|2005
Bucuresti|Inst.Mises Romania|2005
Bucuresti|ALL|2001
Bucuresti|Humanitas|2004
```

Cred că este foarte important să facem experimente cu asemenea comenzi. Puteți experimenta, de pildă, cu condițiile căutării.

Căutarea descrisă anterior era necondiționată. La un volum mare de date asemenea căutări mai mult ne-ar încurca. Să încercăm să găsim autorii și titlurile pentru cărțile care costă mai puțin de 20 de RON.

```
sqlite> SELECT aut, titlu
...> FROM carti
...> WHERE pret<20;
Forta|SQL
Rothbard|Banii
Nola|Diavolul
```

Condiția este pusă după cuvântul-cheie *WHERE*. Dacă am pune *pret>20* aceasta ar însemna că valoarea din câmpul *pret* este mai mare

5. SQL din perspectiva fluxului informațiilor

de 20. Semnificația operatorului `=` este desigur evidentă. Operatorul `!=` servește la negarea egalității.

O construcție interesantă este *BETWEEN ... AND ...* (cuvinte care înseamnă în engleză *între ... și ...*). Efectul ei poate fi văzut mai jos:

```
sqlite> SELECT aut, titlu
...> FROM carti
...> WHERE pret BETWEEN 17 AND 21;
Forta|SQL
Foucault|A supraveghea
Rothbard|Banii
```

5.5 Ordonarea informațiilor

Închipuiți-vă că titlurile din baza de date reprezintă cărțile achiziționate de o bibliotecă. Ele sunt înregistrate pe măsură ce sunt cumpărate. Pentru a regăsi informația sau pentru a construi un catalog al bibliotecii ar trebui sortate informațiile.

Să examinăm titlurile cărților din baza de date:

```
SELECT titlu FROM carti;
SQL
A supraveghea
Banii
Diavolul
Cercetari
```

Nu sunt evident în ordine alfabetică. Să le ordonăm alfabetic!

```
sqlite> SELECT titlu FROM carti
...> ORDER BY titlu;
A supraveghea
Banii
Cercetari
Diavolul
SQL
```

Putem experimenta ordonări după diverse câmpuri.

Putem ordona, de asemenea, în ordinea dorită; de pildă, putem dispune titlurile în ordinea descreșterii prețului:

5. SQL din perspectiva fluxului informațiilor

```
sqlite> SELECT titlu FROM carti
...> ORDER BY pret DESC;
Cercetari
A supraveghea
Banii
SQL
Diavolul
```

5.6 Prezentarea informației din tabelele SQLite

Înainte de a spune câte ceva despre prezentarea informației din tabelele *sqlite*, ar trebui reamintit un principiu important al învățării oricărui program: află cum ieși din programul respectiv! Ați văzut probabil că *sqlite* are o comandă *.help*, apelând la care obținem scurte descrieri ale comenzilor programului:

```
sqlite bib.db
SQLite version 2.8.16
Enter ".help" for instructions
sqlite> .help
.....
.exit                Exit this program
.header(s) ON|OFF    Turn display of headers on or off
.help                Show this message
.mode MODE           Set mode to one of "line(s)",
                    "column(s)", "insert", "list",
                    or "html"
.....
.width NUM NUM ...   Set column widths for "column" mode
```

Am selectat din mesajele *sqlite* doar pe cele relevante aici. Se vede în orice caz că *.exit* este o comandă de ieșire din program.

În continuare puteți vedea un exemplu de afișare a informației pe coloane:

```
sqlite> .mode column
sqlite> .headers on
sqlite> SELECT aut, titlu FROM carti
...> ORDER BY pret DESC;
aut          titlu
-----
```


5. SQL din perspectiva fluxului informațiilor

Wittgenstein	Cercetari
Foucault	A supraveg
Rothbard	Banii
Forta	SQL
Nola	Diavolul

Informațiile din tabelele *sqlite* pot fi lesne incluse în pagini de web. Mai jos se vede cum *sqlite* produce un fragment de cod *html* (pe care-l putem insera într-un tabel *html*).

```
sqlite> .mode html
sqlite> SELECT aut, titlu FROM carti
...> ORDER BY pret DESC;
<TR><TH>aut</TH><TH>titlu</TH></TR>
<TR><TD>Wittgenstein</TD>
<TD>Cercetari</TD>
</TR>
<TR><TD>Foucault</TD>
<TD>A supraveghea</TD>
</TR>
<TR><TD>Rothbard</TD>
<TD>Banii</TD>
</TR>
<TR><TD>Forta</TD>
<TD>SQL</TD>
</TR>
<TR><TD>Nola</TD>
<TD>Diavolul</TD>
</TR>
```

Codul produs cu ajutorul *sqlite* poate fi integrat manual într-o pagină *html*. Lucrul acesta este util, dar mult mai productivă este integrarea automată a fluxului de date în pagina de *web*.

Includerea informației din tabelele *sqlite* în paginile de web se face cu ajutorul unor programe (scrise în limbajul *php* sau alt limbaj destinat acestui scop). Obținem astfel o pagină de web în care informația este adusă la zi în mod dinamic (fără intervenția mâinii omului).

În figura „Afișarea automată a informațiilor din baza de date“ este vizibil un fragment dintr-un program scris în *php*. Ca și-n cazul când operăm cu comenzi în linie, mai întâi este deschisă baza de date. Dacă operațiunea nu reușește, se emite un mesaj de eroare.

Dacă baza de date a fost deschisă cu succes se extrag datele. Este relativ ușor de recunoscut expresia `SELECT * FROM carti`, care nu este

5. SQL din perspectiva fluxului informațiilor



```
Session Edit View Bookmarks Settings Help
#dl("sqlite.so");
$dbhandle = sqlite_open("bib.db");
if (!$db = sqlite_open("bib.db",0666,&$errmessage)){
    echo "Nu pot deschide baza de date:<br />\n";
    echo $errmessage;
    exit;
}
else {
    $sql= "SELECT * FROM carti";
    $res= sqlite_query($db, $sql);

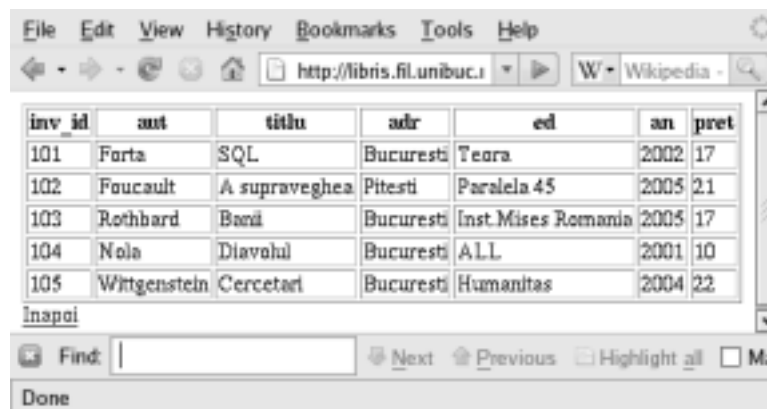
    echo "<TABLE BORDER=1>\n";
    echo "<TR>\n";
                                     2,0-1      5%
```

Figura 5.5: Afișarea automată a informațiilor din baza de date

altceva decât o comandă SQLite care va fi folosită în program. Chiar fără explicarea celorlalte detalii, care ar cere o prezentare a limbajului *php*, este lesne de înțeles ce se-ntâmplă.

Este foarte semnificativ faptul că limbajul *html* nu este suficient. Cu *html* s-ar realiza doar o prezentare statică a informațiilor, acest limbaj nefiind util decât pentru a preciza modul în care informația este afișată pe ecran. Pentru o prezentare dinamică este nevoie de baza de date și de un limbaj de programare care să mențină continuu fluxul informațiilor de la baza de date la pagina de *web*.

Rezultatul obținut este vizibil în figura „Tabel cu cărți pe Internet“.



inv_id	aut	titlu	adr	ed	an	pret
101	Forta	SQL	Bucuresti	Teora	2002	17
102	Foucault	A supraveghea	Pitesti	Paralela 45	2005	21
103	Rothbard	Bană	Bucuresti	Inst.Mises Romania	2005	17
104	Nola	Diavolul	Bucuresti	ALL	2001	10
105	Wittgenstein	Cercetari	Bucuresti	Humanitas	2004	22

Figura 5.6: Tabel cu cărți pe Internet

Ați auzit poate de prescurtarea „LAMP“. „L“ provine de la numele sistemului de operare GNU/Linux, un sistem folosit pe multe com-

5. SQL din perspectiva fluxului informațiilor

putere. „A“ provine de la numele programului care „servește“ clienții cu pagini de *web*; programul se numește „Apache“. „M“ provine de la „MySQL“, programul pentru baza de date. „P“ vine de la „php“, limbajul de programare despre care a fost vorba mai sus. Combinația aceasta este (una dintre cele) folosite pentru a crea pagini de *web* în care fluxul datelor nu este înghețat undeva la un stadiu al său, ci este dinamic.

LAMP sau alte soluții mai noi de acest gen suplimentează capacitățile limitate ale unui sistem de marcare a textului cu o bază de date și un limbaj în care pot fi scrise programe. De îndată ce datele se modifică programele rescriu paginile de *web*, aducând constant la zi informațiile. Mai mult decât atât programele în *php* sau un alt limbaj pot prelua informații trimise de către persoanele care văd paginile de *web*.

5.7 SQL din perspectiva fluxului informațiilor: termeni-cheie

- SQL
- SQLite
- bază de date
- căutare în baza de date
- sortarea (informațiilor extrase din baza de date)
- prezentarea informațiilor (din baza de date)
- automatizarea prezentării informațiilor pe Internet

5.8 SQL din perspectiva fluxului informațiilor: exerciții și analize de caz

Exercițiul 1: Dacă, la firmă sau chiar acasă, s-a cumpărat un program pentru baze de date încercați să vedeți dacă și în ce mod pot fi introduse comenzi în limbajul SQL.

Exercițiul 2: Dacă aveți posibilitatea, descărcați de pe Internet complexul de programe OpenOffice2 <<http://www.openoffice.org>>. Instalați cel puțin versiunea 2.2 pentru a avea și baza de date. Creați o bază de date. Creați o „*Query*” (o interogare a bazei de date) folosind limbajul SQL. Atenție la faptul că numele câmpurilor sau ale tabelelor trebuie puse în ghilimele în OpenOffice2.

Exercițiul 3: Descărcați de pe Internet SQLite. Folosiți ca punct de plecare pagina proiectului SQLite <<http://www.sqlite.org/>>. Alegeți versiunea care vi se pare potrivită. Citiți instrucțiunile de instalare. Instalați SQLite.

Exercițiul 4: Citiți capitolul al 3-lea din cartea lui Ewald Geschwinde și Hans-Jürgen Schönig *PostgreSQL* (București: Teora, 2003). În engleză cartea se numește *PostgreSQL Developer's Handbook*, iar capitolul al 3-lea, „An Introduction to SQL”, este disponibil pe Internet la adresa <http://www.perlhelp.com/sams/Sams_Chapters/0672322609/>. Comparați limbajul SQL înțeles de PostgreSQL cu cel utilizat de SQLite (folosiți pagina <<http://www.sqlite.org/lang.html>>, care descrie limbajul SQL folosit de SQLite).

Exercițiul 5: Refaceți exemplele din această unitate de învățare folosind SQLite instalată la exercițiul 3 și cunoștințele completate cu

5. SQL din perspectiva fluxului informațiilor

ajutorul exercițiului 4. Adaptați eventual exemplele la domeniul care vă interesează în mod special pe dumneavoastră, folosind altceva decât informații despre cărți.

Analiza de caz 1: Multă lume scrie bibliografiile în fișiere de tip *doc* sau similare. Acestea sunt fișiere pentru texte tehnoredactate. Este aceasta o eroare? Gândiți-vă la ce se-ntâmplă dacă bibliografia trebuie prezentată diferit sau dacă trebuie resortată etc. Pentru a documenta răspunsul citiți din Wikipedia articole din categoria „Reference Management Software“ <http://en.wikipedia.org/wiki/Category:Reference_management_software>. Încercați să vedeți ce sistem de management al informațiilor bibliografice s-ar potrivi cu sistemul de scriere computerizată folosit de dumneavoastră.

Analiza de caz 2: Să zicem că o firmă adună informații despre propriul personal, dar le stochează sub formă de fișiere *doc*. Nu le pune într-o bază de date. Dacă informațiile se schimbă, personalul trebuie să rescrie documentele, introducând informațiile aduse la zi. Stabiliți dezavantajele acestui mod de a proceda, prin comparație cu stocarea informațiilor într-o bază de date. Dacă socotiți că există cazuri reale în care se procedează în acest mod, încercați să identificați motivele pentru care managerii firmei procedează în acest fel.

UNITATEA DE ÎNVĂȚARE 6

LaTeX ca mod de organizare a scrierii unui text

„Complexul paginii albe“: terminologia amintește de vremurile când se scria cu creionul pe hârtie, dar fenomenul rămâne real. Mulți oameni nu pot depăși momentul de blocaj generat „pagina“ pe care nu se află nimic. Eroarea este una destul de simplă; ține de iluzia că textul este undeva „în minte“ și trebuie doar transferat de acolo pe hârtie. Textul trebuie construit.

Pentru managerul care are informațiile și are idei cu privire la modul în care vrea să le comunice scrierea unui text n-are de ce să fie o problemă. Ca și-n celelalte unități de învățare a managementului informații vom susține că uneltele din computer pot fi de mare ajutor. Nu vom încerca să vindem însă (obișnuitele) iluzii că un program cu butoane și multe meniuri este soluția cea mai bună. Singura soluție cu adevărat bună este să „vorbești limba“ uneltelor respective. Pilula este amară, dar până la urmă este singurul leac efectiv.

Obiectivele acestei unități de învățare sunt:

- Explicarea utilității separării redactării de tehnoredactare;
- Descrierea beneficiilor limbajului LaTeX, care separă clar crearea conținutului textului de tehnoredactarea textului;
- Explicarea obstacolelor care se ivesc în calea unui mod rațional de organizare a scrierii unui text;
- Înțelegerea etapelor de elaborare a conținutului unui text, cu accent pe transmiterea adecvată a informațiilor și pe mijloacele care facilitează accesul cititoarelor/cititorilor la informațiile din text;
- Precizarea sensurilor în care putem vorbi despre „limbajul tehnoredactării“.

6.1 Sirenele din computer

În *Odiseea*, Ulise și tovarășii săi trebuie să înfrunte pericolul de a se lăsa seduși de sirene. Sirenele sunt ființe ale căror cântece vrăjesc pe oricine. Pentru a nu le auzi, oamenii lui Ulise își înfundă urechile, iar Ulise este legat de catarg.

Circe îl previne pe Ulise: „Întâi și-ntâi sosi-vei la sirene, \ Acele care-ademenesc pe oameni, \ Pe toți care s-apropie de ele. \ Oricine merge-aproape fără știre \ Și cântecul sirenelor aude \ ... E dus, nenorocit pe totdeauna ... “ (Homer, *Odiseea* 12.52-59 trad.Murnu; 12.39-44 în textul homeric). Sunt multe sirene în computerele actuale, dar niciunele nu-s mai periculoase decât cele din editoarele de tehnoredactare. Totul poate fi pus sub forma unui caz care merită analizat.

__Redactare vs. tehoredactare__

În edituri există o diviziune clară a muncii între autoare/autori, redactori, corectori și tehoredactori.

==Problema==

Problema creată de editoarele de tehoredactare WYSIWG este că încurajează o confuzie între ceea ce face cine scrie/redactează un text și ceea ce face cine tehoredactează textul.

[Pe măsură ce studiați această unitate de învățare, veți sesiza desigur celelalte elemente care intră în analiza acestui caz. Autorul textului de față este ferm în favoarea separării redactării de tehoredactare.]

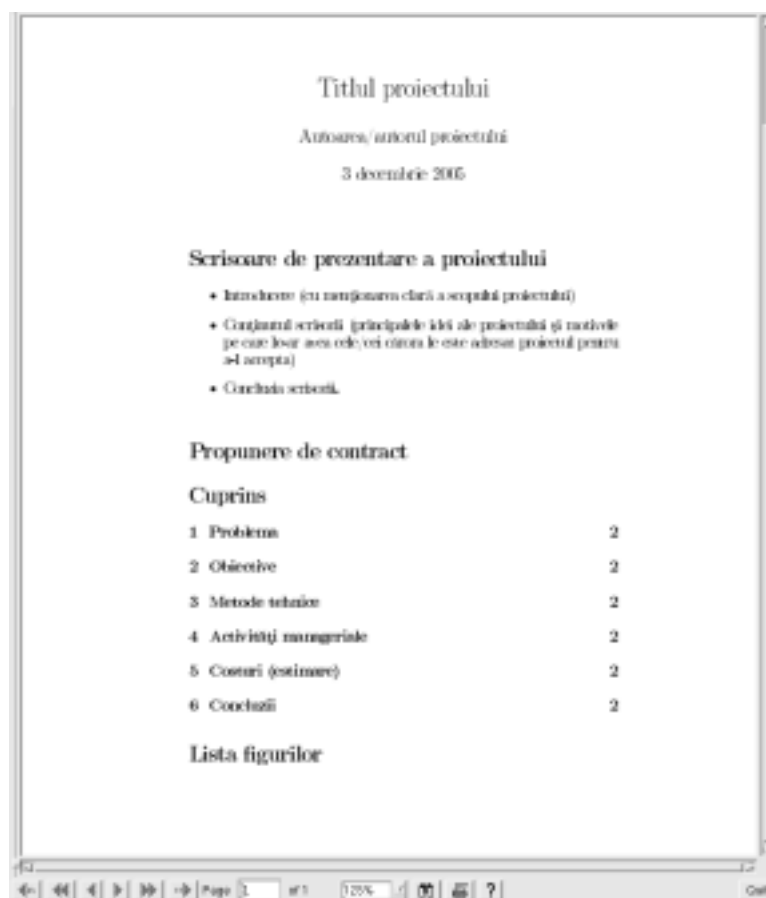
Cântarea sirenei promite ceva foarte atrăgător: tehoredactare simultană cu scrierea textului.¹ Vei vedea pe ecran textul în forma sa

¹Efectele „cântecului sirenelor“ sunt considerabile. Încă de acum un deceniu se vorbea despre dispariția tehoredactorilor. Peter Brielmaier și Eberhard Wolf scriau: „La ora actuală, tot mai multe edituri din Germania încredințează realizarea de pagini întregi redacțiilor, și nu unui corp specializat, tehoredactorii.“ (*Ghid de tehoredactare* [Iași: Polirom, 1999], p.8) Dar și ei adaugă: „Pericolele acestei evoluții sunt evidente: redactorii au tot mai puțin timp pentru conținutul știrilor, iar producția trece din ce în ce mai mult pe primul plan.“ (*ibidem*) Oricum, cei doi autori prezintă totuși o *meserie* distinctă de cea de redactor. Ceea ce nu se relevă, după mine, în cartea lor este faptul că decizia ca o persoană să practice două meserii concomitent este una de management și nu una impusă de tehnologie. De asemenea, Brielmaier și Wolf sunt interesați de tehoredactarea ziarelor. Aici avem în vedere articolele, proiectele, cărțile și nu ziarele.

6. LaTeX ca mod de organizare a scrierii unui text

aranjată pentru tipărire. Punem în paranteză aici discuția cu privire la posibilitatea de a realiza așa ceva din punct de vedere tehnologic. Marele dezavantaj, din perspectiva construirii textului, este că încurajează confuzia dintre conținut și forma tipărită a textului. Iar o formă frumoasă a tiparului a exercitat întotdeauna o tiranie asupra minții oamenilor, care tind să fetișizeze cuvântul tipărit.

Să încercăm să ne gândim puțin la un caz ceva mai concret. Să spunem că avem de condus o echipă care trebuie să prezinte unei comisii un proiect. Este foarte ușor să ne lăsăm fascinați de forme precum cele din figurile „Cuprinsul unui proiect“ sau „Fragmente din conținutul unui proiect“.



Titlul proiectului	
Autorul/autorii proiectului	
3 decembrie 2005	
Scrisoare de prezentare a proiectului	
• Introducere (cu menționarea clară a scopului proiectului)	
• Conținutul scrisorii (principalele idei ale proiectului și motivele pe care le-ar avea cei care îl prezintă, și de ce este necesar proiectul pentru a-l accepta)	
• Concluzia scrisorii.	
Propunere de contract	
Cuprins	
1	Problema 2
2	Obiective 2
3	Metode tehnice 2
4	Activități manageriale 2
5	Costuri (estimare) 2
6	Concluzii 2
Lista figurilor	

Figura 6.1: Cuprinsul unui proiect

Conținutul exemplului din Figura „Cuprinsul unui proiect“ reprezintă o adaptare a schemei pentru un proiect, prezentat ca o propunere, din Raman și Sharma[9, p.462]. Când lucrăm direct în editorul de tehnoredactare textul, afișat frumos pe ecran, începe de să ne fure.

6. LaTeX ca mod de organizare a scrierii unui text

Uităm că sarcina principală, în prima fază, este să gândim conținutul, nu modul cum va arăta textul tehnoredactat.

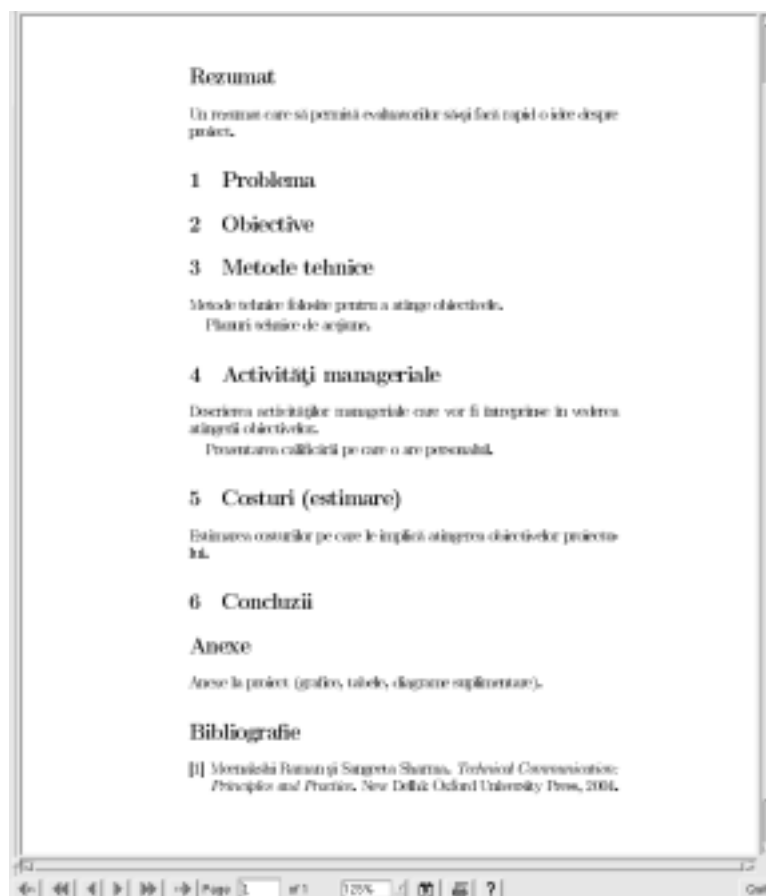


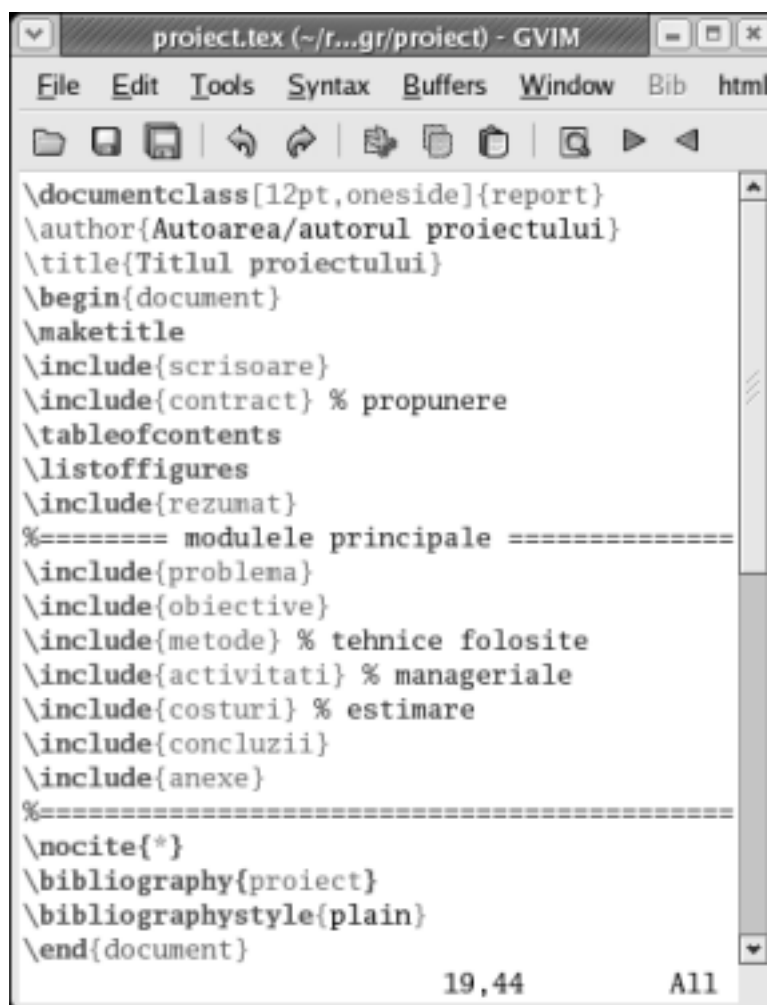
Figura 6.2: Fragmente din conținutul unui proiect

Tehnoredactarea, dincolo de simpla așezare în pagină a textului implică o serie de alte operații. Activitatea de tehnoredactare este departe de a fi simplă și nu întâmplător în edituri există o diviziune a muncii care implică prezența unor tehnoredactori specializați.

6.1.1 Evitarea sirenelor

În *Odissea* oamenii lui Ulise își înfundă urechile. Tot așa ar fi recomandat să procedăm și atunci când scriem un text. O structură precum cea din figura „Proiect în LaTeX“ nu arată deloc atractiv, dar ne fereste de farmecele textului tipărit. Ne forțează practic să ne concentrăm asupra conținutului.

6. LaTeX ca mod de organizare a scrierii unui text

The image shows a screenshot of a Gvim editor window titled "proiect.tex (~/.r...gr/proiect) - GVIM". The window displays a LaTeX source file with the following content:

```
\documentclass[12pt,oneside]{report}
\author{Autoarea/autorul proiectului}
\title{Titlul proiectului}
\begin{document}
\maketitle
\include{scrisoare}
\include{contract} % propunere
\tableofcontents
\listoffigures
\include{rezumat}
%===== modulele principale =====
\include{problema}
\include{obiective}
\include{metode} % tehnice folosite
\include{activitati} % manageriale
\include{costuri} % estimare
\include{concluzii}
\include{anexe}
%=====
\nocite{*}
\bibliography{proiect}
\bibliographystyle{plain}
\end{document}
```

The status bar at the bottom right of the window shows "19,44" and "All".

Figura 6.3: Proiect în LaTeX

Soluția descrisă în continuare folosește LaTeX.² LaTeX este un limbaj de programare orientat către tehnoredactarea textelor. Orientarea către tehnoredactare se vede și din *logo*-ul meșteșugit al LaTeX-ului:

LaTeX este un limbaj de programare în toată puterea cuvântului. Este, tehnic vorbind, universal. Atâta doar că este foarte comod când este vorba de tehnoredactare (și n-ar fi recomandabil pentru scrierea unor programe de calcul numeric, de pildă). Fiind un *limbaj* de programare n-are nici un sens să fie comparat cu un *editor de tehnoredactare*.³ Avantajul său față de editoarele de tehnoredactare este acela că per-

²Pentru mai multe informații despre LaTeX v. *Eseul filosofic*[11].

³Editorul de tehnoredactare OpenOffice, de exemplu, are capacitatea de a genera cod LaTeX.



Figura 6.4: Logo-ul LaTeX

mite o excelentă separare între redactarea textului și tehnoredactarea sa.

Un program LaTeX poate fi scris cu orice editor de texte. Putem folosi, de exemplu, Vim. Există și editoare specializate pentru LaTeX.⁴

Deși termenul „program LaTeX“ intimidează, în realitate scrierea unui astfel de program se învață relativ repede. Există o comunitate puternică în jurul LaTeX și, de fapt, oricine utilizează LaTeX se poate limita la scrierea textului propriu-zis. Restul programelor de care are nevoie sunt deja scrise și sunt disponibile pe Internet.

6.1.2 Problema efectelor de rețea

Poate și mai tare intimidează faptul că LaTeX este un limbaj de programare în care se scriu surse care sunt *compile*. Pe înțelesul tuturor, asta înseamnă că scrii textul, apelezi o unealtă numită „compilator“, iar aceasta furnizează un rezultat... dacă n-ai nici o eroare în program. De aceea, prin definiție, în cazul LaTeX, nu poate fi vorba despre tehnoredactare simultană.

Problema de mai sus nu este însă chiar așa de teribilă dacă scrii textul „pas cu pas“. Pe măsură ce elaborezi o parte din el, compilezi și, dacă este nevoie, elimini erorile pentru a obține un rezultat. Ar fi mai greu dacă ai scrie întâi o carte de 200 de pagini și, la urmă, ai încerca să compilezi. Dacă sunt multe erori, va fi extrem de greu, chiar pentru cineva cu experiență, să le elimine.

Adevărata problemă cu compilatoarele de LaTeX este însă alta - una ce ține mai degrabă de management, decât de opțiunile tehnologice. Fișierele produse de compilatoarele LaTeX sunt de obicei de trei tipuri *dvi*, *ps* și *pdf*. Primul este un format „intern“. Al doilea format este PostScript. PostScript este, de fapt, tot un limbaj, dar al printerelor. Așa cum ne-am aștepta, dacă un compilator uzual traduce sursa în limbaj-mașină (limbajul înțeles de procesorul computerului), uneltele de compilare LaTeX fac tot o traducere într-un limbaj „pe înțelesul“ unei mașini, dar al printerului. Pdf este înrudit cu PostScript, dar este destinat tipăririi pe o hârtie virtuală. Altfel spus, pdf-urile pot să fie citite atât pe ecran, cât și pe hârtie.

⁴A se vedea explicațiile din *Eseul filosofic*[11], unde este prezentat TeXnicCenter.

6. LaTeX ca mod de organizare a scrierii unui text

Oricine a citit alineatul anterior și are fie și o idee, fie chiar și una foarte vagă, despre computere vede că lipsește formatul de fișier *doc*. Acest format de fișier este cerut în multe locuri. De pildă, am tradus recent o carte. Editura la care am dus textul a vrut un fișier *doc*.

Cei mai mulți colegi sau colaboratori externi de la Universitatea unde lucrez folosesc formatul *doc* pentru texte electronice. S-ar putea ca mie să nu-mi convină acest lucru, dar apare *efectul de rețea*: dacă vreau să particip la cercetări sau chiar doar să lucrez în continuare în același loc de muncă, trebuie să folosesc, pentru a coopera, formatul *doc*.

Cazul care apare este unul cât se poate de interesant. Pe de o parte sunt beneficiile cooperării cu persoane care folosesc formatul *doc*. Pe de altă parte sunt beneficiile unui sistem precum LaTeX, care-ți permite să-ți organizezi minunat scrisul, dar nu produce (în mod direct) fișiere *doc*.

Pare cvadratura cercului. O vreme am scris unele texte în LaTeX și altele cu OpenOffice, care este capabil să producă și fișiere *doc*. Am încercat să folosesc și un convertor din LaTeX în rtf, un limbaj de marcarea a textului pe care-l înțeleg cam toate editoarele de tehnoredactare (inclusiv Word), dar m-am lovit de o problemă: codurile literelor din sistemul de operare cu care lucrez nu sunt înțelese de convertor. Părea să nu existe nici o soluție.

Rezolvarea am găsit-o într-o zi, când am gândit lucrurile din perspectivă mai degrabă managerială decât tehnologică. Cele mai multe dintre documentele pe care le folosesc în procesul de cooperare descris mai sus folosesc doar o parte (destul de mică) a limbajului LaTeX. Mi-ar fi imposibil să scriu de unul singur un convertor pentru *întregul* limbaj LaTeX. S-a dovedit însă mult mai simplă scrierea unui program care traduce doar o parte din ceea ce se poate scrie în LaTeX.⁵ Traducerea se poate face fie în *rtf*, fie în *html*.

De ce *rtf*? De ce nu *doc*? Răspunsul este cât se poate de simplu. Formatul *rtf* este specificat în documente accesibile public.⁶ Prin contrast, formatul *doc*, cel puțin versiunile sale tradiționale, nu este public.

⁵Programele sunt scrise într-un limbaj numit Python. Soluția a implicat deci iarăși folosirea unui limbaj de programare! Cursul de față, de pildă, folosește acest set de programe. După ce le mai testez o vreme am să le pun și pe pagina mea de web. Utilizarea lor nu va fi însă niciodată o chestiune simplă sau lipsită de riscuri. Oricine ar fi curioasă să vadă cum merg ar trebui să înțeleagă cât de cât Python și să ia măsurile de precauție necesare pentru a nu-și distruge accidental fișierele LaTeX.

⁶Există și o carte despre *rtf*: Sean M. Burke, *RTF Pocket Guide* (O'Reilly, 2003). O parte extrem de utilă a cărții, „RTF Tutorial“ este accesibilă pe Internet <<http://www.oreilly.com/catalog/rtfpg/chapter/part1.pdf>>.

6. LaTeX ca mod de organizare a scrierii unui text

Cu alte cuvinte, doar firma proprietară și contractanții ei știu exact ce trebuie pus, unde și cum într-un fișier *doc*.

Trebuie precizat însă un lucru: ceea ce transmiți cu ajutorul formatului *rtf* este conținutul textului, inclusiv sublinierile, secțiunile, notele de subsol, imaginile ș.a.m.d. Tehnoredactarea, ca să zic așa, este refăcută. Fie este refăcută automat de programul cu care este vizualizat *rtf*-ul, fie, în cazul unei edituri, de personalul din sectorul de tehnoredactare. Dar tocmai acesta este și scopul urmărit. Dacă vreau să trimit textul tehnoredactat deja, atunci un *pdf* este mijlocul potrivit.

Este efectul de rețea un mit? Dacă este văzut prin prisma efectelor sale presupus negative, da - este un mit. Orice piață prinde participante în rețeaua cooperării. În sensul acesta există o „dependență“ a individului, dar ea merită menținută cât timp aduce beneficii. Indivizii sunt independenți tocmai pentru că, atunci când sesizează dezavantaje le elimină născocind ceva.

Morala acestui caz trece mult dincolo limitele sale destul de restrânse. În primul rând, orice firmă merită să se gândească dacă merită să mizeze exclusiv pe programe cumpărate sau are sens să angajeze pe cineva care să scrie programe, să „meșterească“ legături, completări la programele achiziționate, în funcție de contextul concret al problemelor firmei. Iar nu în ultimul rând, ideea este că merită investit în cunoaștere. Altfel firma este penalizată de (mitice) relații de dependență.

6.2 A scrie cu gândul la text

Ceea ce prezentăm în această secțiune este un mod de a scrie și nu trăsăturile limbajului LaTeX. Pentru a deprinde LaTeX trebuie să citiți un manual specializat.

Principalul atu al LaTeX este separarea realizării conținutului textului de tehnoredactarea acestuia. Acest avantaj este deosebit de important în special în cazul textului lucrărilor științifice, a rapoartelor de cercetare etc.

6.2.1 Clarificarea obiectivelor

Există un leac excelent pentru „blocajul în fața foii albe“: să scriem un text -care nu va face parte din textul final - în care enunțăm obiectivele pe care le urmărim.⁷ Acest text poate să fie, tehnologic vorbind,

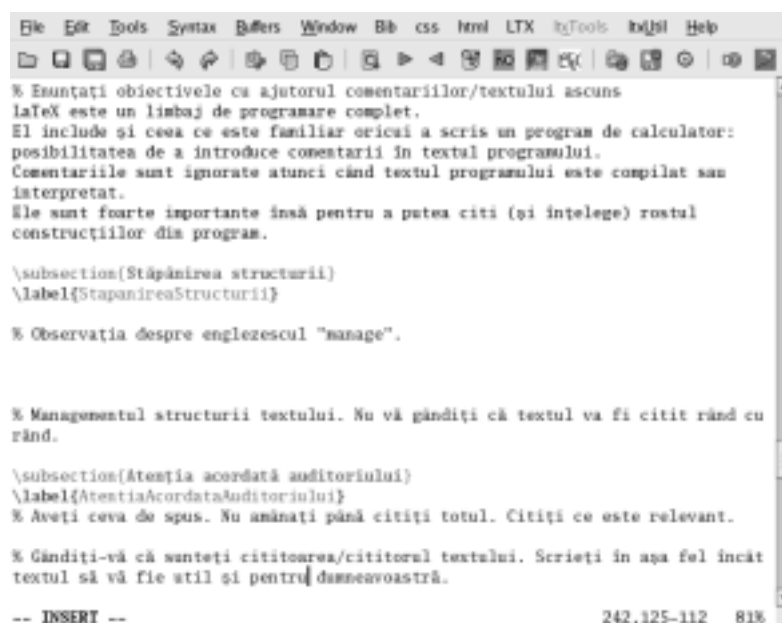
⁷Conducătorul tezei mele de doctorat, când a simțit că nu reușesc să aștern pe hârtie textul final al tezei, mi-a spus o anecdotă pe care n-am uitat-o niciodată: în ea era vorba despre un doctorand care-și începea teza cu „scopul acestei lucrări de

6. LaTeX ca mod de organizare a scrierii unui text

complet separat (un fișier separat) sau să fie ascuns în sursa textului, ca un comentariu de genul celor pe care le folosesc programatorii.

LaTeX este un limbaj de programare complet. El include și ceea ce este familiar oricui a scris un program de calculator: posibilitatea de a introduce comentarii în textul programului. Comentariile sunt ignorate atunci când textul programului este compilat sau interpretat. Ele sunt foarte importante însă pentru a putea citi (și înțelege) rostul construcțiilor din program.

În LaTeX comentariile sunt rânduri sau porțiuni de rânduri care încep cu semnul procentului. După cum se vede din figura „Enunțarea obiectivelor textului“, când am scris unitatea de învățare mi-am notat chiar acest obiectiv: să scriu despre enunțarea scopului urmărit în porțiunea respectivă de text.





```
File Edit Tools Syntax Buffers Window Bib. css html LTX  Tools  Help  
% Enunțați obiectivele cu ajutorul comentariilor/textului ascuns  
LaTeX este un limbaj de programare complet.  
El include și ceea ce este familiar oricui a scris un program de calculator:  
posibilitatea de a introduce comentarii în textul programului.  
Comentariile sunt ignorate atunci când textul programului este compilat sau  
interpretat.  
Ele sunt foarte importante însă pentru a putea citi (și înțelege) rostul  
construcțiilor din program.  
  
\subsection{Stăpânirea structurii}  
\label{StapanireaStructurii}  
  
% Observația despre englezescul "manage".  
  
  
  
% Managementul structurii textului. Nu vă gândiți că textul va fi citit rând cu  
rând.  
  
  
  
\subsection{Atenția acordată auditoriului}  
\label{AtentiaAcordataAuditoriului}  
% Aveți ceva de spus. Nu amânați până citiți totul. Citiți ce este relevant.  
  
% Gândiți-vă că sunteți cititorul/cititoarea textului. Scrieți în așa fel încât  
textul să vă fie util și pentru dumneavoastră.  
  
-- INSERT -- 242,125-112 81%
```

Figura 6.5: Enunțarea obiectivelor textului

În Vim textul comentariilor are o culoare distinctă, pentru a evita confuzia cu restul textului. După compilarea cu LaTeX comentariile dispar. Ele nu vor apărea la tipărirea textului tehnoredactat.

doctorat este obținerea titlului de doctor“. Sensul adânc al glumei este că nu trebuie să încerci să obții alte rezultate decât acelea care sunt la nivelul necesar obținerii titlului de doctor. Dacă-ți propui „să faci o mare descoperire“, lucrul acesta este o eroare. Poți să încerci asta în lucrări ulterioare, dar scopul tezei este doar să demonstreze că ai aptitudini de cercetătoare/cercetător. Desigur, doctorandul n-ar fi trebuit să-și enunțe obiectivul în textul „vizibil“.

6. LaTeX ca mod de organizare a scrierii unui text

Este bine desigur să începem cu enunțarea obiectivelor mari ale textului. Le putem plasa undeva la început. Dacă există o coincidență cu obiectivul major al unui capitol, vom pune comentariul la începutul capitolului respectiv.

Uneori se mai întâmplă să ne vină câte o idee, dar n-avem timp, în acel moment, să-i dăm forma finală. O plasăm ca un comentariu în text, la locul potrivit. Ulterior vom reveni și vom finaliza ideea.

Comentariile sunt foarte utile și atunci când recitim textul. Acolo unde credem că trebuie refăcut ceva punem un comentariu. În LaTeX există chiar un „pachet“ (un modul apelat la cerere) care permite inserarea unor atenționări în text. La compilare ni se va atrage atenția asupra locului în care mai este ceva de reparat în text. Pe măsură ce remediem textul scoatem și semnalele de atenționare.

6.2.2 Stăpânirea structurii

Ca orice neologism, „management“ este un cuvânt stingher în limba română: nu prea are rude, prieteni. Lucru și mai important: nu există în familie un verb, precum *to manage* din engleză. Cu *manage* poți spune în engleză multe lucruri: conduci, ai grijă, te descurci, atingi un scop.⁸

În cazul unui text, ceva trebuie să „*manage*“ în mod cert: structura textului. Există două motivații ale acestui management al structurii textului. Capitolele, secțiunile, subsecțiunile sunt utile pentru cine scrie. Pentru cine scrie, structurile textului sunt precum niște module pe care le combini pentru realiza construcția finală. Pentru cine citește, structurile permit o viziune panoramică asupra textului, îi permit „să-și facă o idee“ despre ce se spune în text.

Eroarea majoră pe care o comit mulți este aceea de a proiecta structurile înaintea obiectivelor. Obiectivele implică, de multe ori și o cercetare; în orice caz implică o evaluare a informațiilor pe care vrem să le comunicăm (pentru a atinge obiectivele propuse). Cu acest prilej îți faci o imagine despre ideile-cheie ale viitorului text și abia apoi poți proiecta un element al structurii care să pivoteze în jurul cheii găsite.

În LaTeX există comenzi pentru capitolele, secțiunile, subsecțiunile textului. De pildă, `\section{Titlul secțiunii}` îi spune sistemului că începe o nouă secțiune.

Comenzile LaTeX de secționare a textului sunt o bună ilustrare a ideii de focalizare pe conținutul, nu pe tehnoredactarea textului. Nu spunem cât de mare să fie litera cu care este scris titlul sau ce decora-

⁸Wordnet a dat o listă cu șapte sensuri ale lui *manage* <<http://wordnet.princeton.edu/perl/webwn>>.

6. LaTeX ca mod de organizare a scrierii unui text

țiuni să fie plasate în jurul său. Nu vom indica nici care ține de așezarea finală în pagină.

Putem deplasa cu ușurință secțiunile de text sau alte elemente ale structurii. Pur și simplu le tăiem și le copiem în altă parte (evident cu comanda corespunzătoare cu tot). Nu trebuie să ne facem griji cu privire la numerele secțiunilor, cuprins ș.a.m.d. Totul se va schimba automat.

6.2.3 Atenția acordată auditoriului

Atunci când scrieți gândiți-vă că aveți ceva de spus. Ar fi inutil să amânați redactarea textului până citiți tot ce s-a spus pe tema respectivă. Are sens să citim doar ceea ce este relevant. Chiar și așa, există un punct de oprire: e momentul în care avem suficiente informații pentru a putea regândi problema în discuție cu propria minte.

Există două momente între care trebuie găsit apoi un echilibru. Pe de o parte, este firesc (și extrem de util) să vă gândiți că sunteți cititoarea sau cititorul textului pe care-l scrieți. Puneți-vă „în pielea lor“. Dacă scrieți o frază lungă și încâlcită, gândiți-vă ce ați pricepe din ea dacă ați fi în postura de cititoare/cititor. Desigur, puțin. Refaceți fraza. Dacă presupuneți că un anumit lucru este „bine cunoscut“, încercați să vedeți dacă așa gândesc și alții.⁹

Pe de altă parte, este un bun sfat să scrieți în așa fel încât ceea ce puneți pe hârtie să vă fie util și dumneavoastră. De pildă, am scris într-una din unitățile de învățare despre datele calendaristice. Am menționat și standardul ISO relevant. Este greu de crezut că, peste ceva timp, voi mai ține minte numărul și locul unde pot găsi acel standard. Pot să revin însă la propriul text și să găsesc informațiile de care am nevoie. Atâta timp cât respect această regulă de a pune în text informații care-mi folosesc și mie ca autoare/autor există o bună șansă ca orice altă persoană, aflată în aceeași situație, să aibă și ea nevoie de respectivele informații.

Nu este ușor de găsit un echilibru între aceste două cerințe. Pe de o parte poți exagera punând doar ce ți-ar fi ție util. Pe de altă parte, nu este foarte simplu să-ți dai seama ce le poate fi de folos altora. Nu există însă reguli pentru a atinge echilibrul. Cel mai simplu este să scrii doar ceea ce ai priceput cu propria minte și să încerci să „citești“ mintea auditoriului ca să vezi asupra căror informații are rost să insiști.

⁹Chestiunea aceasta este adevărată și în sens invers. Când citiți aceasta unitate de învățare nu lucrați cu presupunerea „toată lumea scrie cu Word“. Eu unul, de pildă, nu scriu cu Word. Mai grav ar fi însă să vă gândiți că „scriu cu LaTeX“. Acest gând n-are noimă: LaTeX este un limbaj, nu un editor. Pot să scriu *în* limbajul LaTeX.

6. LaTeX ca mod de organizare a scrierii unui text

Sfaturile acestea sunt evidente, dar adesea nu sunt respectate. Este evident că o matematiciană care scrie pentru publicul de specialitate n-are rost să insiste asupra chestiunilor elementare. Studentele/studentii care scriu lucrări ce vor fi citite de profesoare/profesori procedează însă uneori în același mod. Nu le mai spun lucruri pe care aceștia le știu deja prea bine. În acest caz, auditoriul a fost evaluat greșit. Profesorul/profesoara, când citesc lucrările studentelor/studentilor, sunt *examinatoare/examinatori*. Tocmai de aceea este bine să le arăți *că știi* ceea ce știu și ei.

Altă eroare a studentelor/studentilor este aceea de a le spune profesoarelor/profesorilor *numai* ceea ce știu și aceștia. Examinatoarele/examinatorii vor să vadă și dacă poți judeca folosind mintea proprie. Se așteaptă nu să le arăți doar că aveți un teren comun pentru discuție, ci și o minte care gândește independent.¹⁰

6.2.4 Cărămizile textului

Uneori textul este comparat cu o țesătură. Comentariile invizibile sunt precum urzeala. Printre firele de urzeală le petrecem pe cele de bătătură. Propozițiile, frazele ar fi bătătura.

Alteori comparația este cu o construcție. Comentariile care enunță obiectivele sunt asemenea proiectului constructorilor. Structuri precum capitolele și secțiunile seamănă cu scheletul de beton al unei clădiri. Alineatele sunt precum pereții. Propozițiile sunt adoma cărămizilor. Imaginile sunt ferestre. Trimiterile sunt precum cablurile sau conductele ce leagă între ele instalațiile.

Limbaajul LaTeX poate fi de mare folos pentru a țese/a construi structurile de amănunt ale textului: propozițiile, frazele, alineatele. Pentru aceasta este absolut recomandabil, atunci când limbaajul ne este cunoscut, să scriem cu un editor de genul Vim-ului.¹¹

Ar trebui spus câteva lucruri și despre marcajele introduse în text cu ajutorul LaTeX. Conținut nu înseamnă numai cuvinte combinate în propoziții. Unele cuvinte, de pildă, vrem să le evidențiem. Alteori vrem să introducem o trimitere ș.a.m.d. Toate aceste lucruri trebuie marcate în text.

Cel mai simplu marcaj LaTeX este o pereche de acolade. Textul pus între acolade va fi tratat într-un mod aparte. Ce anume se întâmplă

¹⁰Discut aici desigur cazul ideal. Ignor profesorul care nu vrea să audă decât ceea ce a spus la curs.

¹¹Pentru începătoare/începători sau persoane care ar scrie mai rar în limbaajul respectiv există editoare speciale excelente. A se vedea, de pildă, prezentarea TeXnicCenter-ului făcută în *Eseul filosofic*[11].

6. LaTeX ca mod de organizare a scrierii unui text

cu textul respectiv poate fi indicat, de pildă, punând un cuvânt în fața acoladelor. Pentru ca acest cuvânt să nu se confunde cu restul textului el este precedat de o bară oblică inversă. De pildă, `\textit{}` spune că textul dintre acolade trebuie scris cu italice.

O altă metodă este să scriem, de exemplu `{\em }`, pentru a arăta că textul din acolade care urmează după `\em` trebuie evidențiat. Metoda de evidențiere este aleasă la tehnoredactare. În conținut specificăm doar că porțiunea respectivă de text este scoasă în evidență.

LaTeX este un limbaj de programare universal. Nu este deci nici cea mai mică problemă dacă scriem: `\subliniat{}`. Trebuie însă să definim ce anume înțelegem prin această expresie. Ar putea fi, desigur, un sinonim al lui `\underline{}`.

Din punctul de vedere al persoanei care scrie, claritatea marcajelor LaTeX este remarcabilă.¹² Mai mult decât atât, așezarea propozițiilor se face logic, nu vizual.

Fiecare propoziție poate (și este absolut recomandabil) să fie așezată pe alt rând logic. Un rând alb este semnul că s-a încheiat alineatul. Avantajul acestei dispunerii a textului este că astfel pot fi urmărite lesne raporturile dintre propoziții în cadrul alineatului. De exemplu, dacă prima propoziție este cheia alineatului, putem verifica lesne dacă următoarele propoziții explică, susțin sau exemplifică ideea din propoziția-cheie.

Notele „de subsol“ sunt, în sursa LaTeX, paranteze în text. Scriem `\footnote{}` și punem între acolade conținutul notei. Așa este *logic* să procedăm: notele de subsol trebuie gândite ca paranteze în cadrul (conținutului) textului. Punerea lor în subsol (sau la sfârșitul textului) este o chestiune ce ține de tehnoredactare.

6.2.5 Trimiterile

S-a scris mult (și în termeni exagerați) despre noutatea hipertextului. Noutatea rezidă în viteza cu care găsim textul la care se face trimitere. Altfel s-au făcut de multă vreme trimiteri: la bibliografie, la un loc dintr-o carte pe care o găseai în bibliotecă, la alte secțiuni sau pagini ale textului. Pentru a ajunge la trimiterile respective parcurgeai efectiv un drum mai lung sau mai scurt. În cazul hipertextului, un fragment

¹²„Prețul“ îl plătește cine programează în LaTeX. Dincolo de un anumit punct scrierea de programe TeX/LaTeX nu este un lucru tocmai simplu. Lucrul acesta este valabil, tot dincolo de un anumit prag de complexitate, și când încerci să convertești marcajele LaTeX în marcajele folosite în alte limbaje. Cum există însă o sumedenie de programe gata elaborate, cine scrie în LaTeX n-are viața chiar așa de grea cum pare la prima vedere.

6. LaTeX ca mod de organizare a scrierii unui text

poate fi undeva în Europa, altul în Australia și al treilea în Japonia, dar tehnologia diferită permite găsierea extrem de rapidă a legăturilor.

Trimiterile joacă un rol deosebit de important, în logica textului, atunci când ele indică *sursa* informațiilor. Este o mare deosebire între un text care indică sursele și unul care nu le indică. Am răsfoit nu demult, într-un anticariat, o carte despre ultimele luni de existență a URSS. Era vorba acolo despre un moment de cădere în cariera unui politician sovietic. Autorul relatează convorbirea sa cu politicianul aflat în dizgrație și introduce în același timp o serie de informații. Ne spune că respectivul făcuse o mare avere (din peșcheșurile subordonaților) și alte lucruri din biografia acestuia. Nu ne indică însă sursa informațiilor sale. Ce este asta? Jurnalism? Poate. Dar pe cine să credem? Pe autor? Pe politician - care susține cu totul altceva. Fără indicarea surselor totul se transformă într-un *argument al autorității* (autorului). Cu surse, informațiile pot fi evaluate critic.

Cum trebuie indicate sursele? În mod explicit și precis. „În mod explicit“ înseamnă că nu este suficient să spui „autorul este trecut în bibliografie“. Asta este plagiat. LaTeX include o comandă de trimitere explicită la bibliografie: `\cite[]{}{}`. Între acolade se trece o etichetă. Este eticheta înregistrării din baza de date a informațiilor privitoare la articolul, cartea, documentul de pe Internet ș.a.m.d. citate în text. Între parantezele drepte se indică locul din sursă - după caz, acesta poate fi o pagină, o secțiune ș.a.m.d.

Se observă că LaTeX folosește o bază de date pentru a stoca informațiile bibliografice. Lucrul acesta este absolut recomandabil.

Alteori, trimiterile sunt pur interne. Putem trimite de la un capitol la altul. Putem indica o pagină ș.a.m.d.

Cuprinsul este și el o colecție de trimiteri. O altă colecție de trimiteri este indexul unei cărți. Pe vremuri am lucrat ca documentarist. Erau timpurile când Marx era un „clasic“ oficial. Într-o zi a sunat cineva „de sus“. Dorea să afle repede unde a afirmat Marx că Lincoln este un fiu al clasei muncitoare. Închipuți-vă peste treizeci de volume cu scrierile lui Marx, amalgamate cu operele lui Engels. Fără indici căutarea ar fi fost infernală. Lucrând cu ipoteza (plauzibilă) că textul este din perioada când Lincoln era în viață și cu indicii găsierea pasajului respectiv este o problemă soluționabilă rapid.

Astăzi este epoca textelor electronice și a motoarelor de căutare. Am introdus în Google doar aceste cinci cuvinte cheie: `Marx Lincoln son working class`. Rezultatul căutării a fost afișat într-o clipă. Pe primul loc printre documentele găsite a fost plasată „Marx's Letter to Abraham Lincoln“.¹³

¹³Este vorba, de fapt, despre un mesaj, scris de către Marx, și trimis în numele Internaționalei <<http://www.marxists.org/history/international/>

Google nu anulează rolul unui element precum indicele textului sau glosarele cu termeni. Din contră. Alegerea cuvintelor-cheie este o decizie luată de om. De aici și importanța unui sistem care să precizeze, din punctul de vedere al autoarei/autorului, care sunt termenii-cheie. Un asemenea rol îl joacă indicii, glosarele sau listele de termeni.

6.3 Limbajul tehnoredactării

Tehnoredactarea este, în fond, *design*-ul grafic al textului. Este vorba despre text ca atare. *Design*-ul grafic înseamnă mai mult, pentru că este vorba și despre ilustrații.

Literele în sine sunt mici desene. Realizarea lor este nu doar o chestiune ce ține de tehnică. Ține de artă.

Modul de a dispune titlurile, de a combina literele alese ș.a.m.d. presupun un întreg complex de operații. LaTeX are capacități deosebite de a realiza cu succes aceste operații în special când este vorba despre formulele matematice.

Este foarte important să ne gândim la tehnoredactare ca la o artă de a spune ceva. Dacă un titlu este scris cu litere mai mari vrem să atragem atenția asupra sa. Tipul de literă ales poate să sugereze, de pildă, o epocă istorică. Într-un text despre Anglia medievală, litere care amintesc de cele folosite în Evul Mediu englez sunt un mod de a crea o atmosferă.

Indicațiile de tehnoredactare nu spun doar *cum* să fie așezat textul în pagină. Ele și transmit un *mesaj*.

Resursele LaTeX în materie de tehnoredactare sunt considerabile. Problema managerială este însă că, în vreme ce este relativ ușor să traducem conținutul textului în alt limbaj, partea de tehnoredactare făcută cu LaTeX este foarte greu de transferat.¹⁴

Logo-ul LaTeX (vezi figura „Logo-ul LaTeX“) este realizat cu ajutorul limbajului de tehnoredactare. El nu este în LaTeX rodul unei instrucțiuni de introducere a unei imagini, ci un șir de comenzi care indică deplasările literelor. Dacă în alt limbaj n-avem la dispoziție comenzile corespunzătoare, „ \LaTeX “ va arăta altfel decât atunci când folosim limbajul LaTeX.

www.wma.com/documents/1864/lincoln-letter.htm>. Merită să citiți și răspunsul ambasadorului Charles Francis Adams, căruia i-a fost prezentat mesajul. Acesta arată, în esență, că președintele Lincoln urmărește realizarea egalității și dreptății pentru toți oamenii, nu doar pentru „oamenii muncii“.

¹⁴În practică, LaTeX este un standard *de facto* în publicațiile de matematică. Marile edituri academice îl folosesc, în special atunci când textele cuprind multe formule matematice.

6. LaTeX ca mod de organizare a scrierii unui text

În ciuda dificultăților de folosire, LaTeX este o paradigmă a organizării scrierii unui text. S-ar putea ca dificultățile de utilizare să se datoreze chiar faptului că scrierea unui text nu este un lucru așa de ușor. Nu există un mod facil de a scrie un text.

6.4 LaTeX ca mod de organizare a scrierii unui text: termeni-cheie

- Redactare
- Tehnoredactare
- Efecte de rețea
- Obiectivele (urmărite atunci când scriem textul)
- Structura textului
- Marcaje LaTeX (în text)
- Trimiteri

6.5 LaTeX ca mod de organizare a scrierii unui text: exerciții și analize de caz

Analiza de caz 1: Vedeți cu ce programe se scriu textele în firma la care lucrați. Dacă se folosesc anumite editoare de tehnoredactare care sunt motivele alegerii lor? Cum se manifestă efectele de rețea în cazuri familiare dumneavoastră? Pentru a adânci argumentele din analiza de caz folosiți ca punct de plecare (către alte surse de informare) articolul din Wikipedia despre „Network effect“ <http://en.wikipedia.org/wiki/Network_effects>.

Analiza de caz 2: Un caz foarte important este cel al formatului fișierelor de computer. Problema este dacă formatele să fie închise (accesibile doar firmei/firmelor care cunosc specificațiile tehnice ale formatelor) sau deschise (accesibile tuturor). Poziția Uniunii Europene este favorabilă formatelor deschise. Citiți documentul „TAC approval on conclusions and recommendations on open document formats“ <<http://ec.europa.eu/idabc/en/document/2592/5588>>. Firma Microsoft, care are o poziție dominantă pe piața programelor de tip *office* își propune și ea să standardizeze un format deschis. Porniți de la articolul „Office Open XML“ <http://en.wikipedia.org/wiki/Office_Open_XML> pentru a găsi mai multe informații.

6. LaTeX ca mod de organizare a scrierii unui text

Exercițiul 1: Aflați mai multe despre *rtf* pornind de la articolul „Rich Text Format“ din Wikipedia <http://en.wikipedia.org/wiki/Rich_Text_Format>. Gândiți-vă dacă i-ați acordat vreodată acestui limbaj de marcare a textului o semnificație specială? Care au fost motivele pentru o atitudine sau alta?

Exercițiul 2: Limbajul LaTeX este, ca să spunem așa, un etaj peste un alt limbaj, numit „TeX“. Aflați mai multe despre TeX citind din Wikipedia articolul „TeX“ <<http://en.wikipedia.org/wiki/TeX>>. Descoperiți cu ajutorul trimiterilor din acest articol exemple de folosire a TeX. Descoperiți, tot cu ajutorul trimiterilor, cărți despre TeX disponibile liber pe Internet.

Exercițiul 3: Aflați mai multe despre LaTeX citind din Wikipedia articolul „LaTeX“ <<http://en.wikipedia.org/wiki/TeX>>. Descoperiți cu ajutorul trimiterilor din acest articol comparații între Word și LaTeX.

Exercițiul 4: Există un limbaj de marcare a textelor cu care vă întâlniți probabil foarte des și care include comentarii: *html*. În *browser*-ul de *web* identificați posibilitatea de a vedea sursa paginilor scrise în *html*. Identificați textul scris între `<--!` și `-->`. Acestea sunt comentariile incluse în sursa *html*.

Exercițiul 5: Date fiind „efectele de rețea“ este puțin probabil că multă lume, după citirea acestei unități de învățare se va apuca să folosească LaTeX. Multe lucruri care se pot face însă folosind limbajul LaTeX se pot face și cu editoare de tehnoredactare uzuale. Încercați să descoperiți cum puteți introduce text ascuns (*hidden text*). Scrieți comentarii cu ajutorul textului ascuns. Încercați să vedeți cum vă ajuta să realizați mai bine un text.

Exercițiul 6: Editoarele de tehnoredactare au posibilitatea de a structura logic textul. Multă lume scrie însă titlurile de secțiuni etc. ca pe niște texte cu litere „un pic mai mari“. Identificați felul în care editorul secționează textul (poate cu „insert section“ sau ceva analog). Identificați rolul jucat de „heading 1“, „heading 2“ ș.a.m.d. (pe care le găsiți într-un meniu de genul „styles and formatting“). Vedeți dacă editorul de tehnoredactare poate genera o „outline“ (o imagine de ansamblu a structurii textului). Examinați modul în care se poate naviga în text de la o secțiune la alta.

Exercițiul 7: Deschideți Wikipedia și căutați la întâmplare articole. Vedeți care au trimiteri la sursele informațiilor comunicate în text. Ce vă spune lipsa trimiterilor despre calitatea articolului? Este bine să evitați asemenea articole? Dacă, dintr-un motiv sau altul, vă decideți să le folosiți cum ar putea un motor de căutare să vă ajute să dați de surse? Testați practic folosirea motorului de căutare pentru găsirea

6. LaTeX ca mod de organizare a scrierii unui text

unor surse (de pildă, dacă găsiți un citat fără sursă, găsiți-o cu motorul de căutare).

UNITATEA DE ÎNVĂȚARE 7

Etica și managementul informației

Într-o epocă în care informațiile circulă tot mai mult prin intermediul computerului există o strânsă legătură între problemele etice ale managementului informației și problemele etice ale tehnologiei informației. De aceea începem această unitate de învățare cu o trecere în revistă a rădăcinilor problemelor etice în natura tehnologiei informației ca atare și în licențele care însoțesc programele de calculator.

Pe de altă parte, strângerea de informații a implicat întotdeauna grave responsabilități. Înainte de cel de al doilea război mondial o serie de țări europene aveau date foarte precise cu privire la religia cetățenilor, locul lor de domiciliu ș.a.m.d. Naziștii au folosit asemenea date pentru a-i identifica pe evrei în timpul Holocaustului. Există întotdeauna pericolul ca date extrem de bine personalizate să fie exploatare de un regim autoritar în scopuri criminale. Pentru cercetările de demografie ar fi fost însă suficientă păstrarea datelor fără a face legături cu persoane fizice, interesul fiind acela de a realiza o *statistică* de mare acuratețe (a se vedea și „Analiza de caz 1“).

Obiectivele acestei unități de învățare sunt:

- Ajungerea la o înțelegere a fundalului tehnologic;
- Analiza diversității licențelor pentru programele de calculator;
- Explicarea motivelor pentru care managementul informațiilor ne pune în fața unor dificile dileme etice.

7.1 Invizibilitatea tehnologiei informației

Când vorbim despre tehnologie, în mintea noastră apar imediat imagini cu materiale, unelte, mașini, produse. Pentru ca să producem o poliță de lemn, avem nevoie de o scândură și alte materiale, de unelte de tâmplărie ș.a.m.d. Toate acestea sunt *vizibile*, tot așa cum sunt lesne de perceput cu ochiul acțiunile de tăiere a lemnului, de șlefuire sau de găurire.

În schimb, dacă spunem că *grep* este o unealtă, aproape că nu ne putem reprezenta vizual nimic. Sunt niște comenzi pe ecran și atât. Undeva, în măruntaiele computerului, au desigur loc niște procese de ordin fizic, dar ele nu ne sunt direct accesibile. Și nici nu sunt tocmai relevante. Mult mai important ar fi să știm ce algoritmică folosește programul *grep*. Algoritmul este însă un obiect de tip matematic, nu unul fizic.

Discuția despre tehnologia informației suferă adesea din pricină că, pentru cine nu este familiarizat cu ea, e extrem de dificil să-și reprezinte ce se întâmplă de fapt. Se folosesc o mulțime de analogii. Unele sunt foarte importante din perspectivă etică. De pildă, se vorbește despre spărgători de baze de date, de programe etc. Are analogia aceasta o noimă? Și dacă da, atunci care sunt limitele în care ea funcționează.

Când este vorba despre terminologia relevantă etic a tehnologiei informației nici un caz nu este mai grăitor decât al termenilor englezești *hacker* și *cracker*. Problema care se pune în acest caz poate fi sistematizată în felul următor:

`__Meșter(_hacker_) vs. spărgător(_cracker)__`

Oamenii sesizează cu ușurință înțelesurile când pot `_vedea_` despre ce este vorba. Orice lăcătuș bun poate să desfacă încuietoarea unei uși. Nimeni nu folosește termenul „lăcătuș“ ca pe un sinonim pentru „spărgător“.

Lucrurile se schimbă însă când oamenilor le este greu să `_vadă_` despre ce este vorba.

==Problema==

Dacă vă cumpărați o haină și rupe un nasture îl coaseți la loc. Nu anunțați firma producătoare, nu aruncați haina și luați alta. Probabil vi se pare firesc să puteți modifica haina respectivă. Ei bine atunci sunteți un `_hacker_` de haine.

7. Etica și managementul informației

Are vreun sens ca o activitate analoagă în cazul programelor de computer să fie considerată condamnabilă? Sau ar trebui să rezervăm acest condamnările noastre pentru cei care deschid fișiere din computerele altora, modifică funcționarea computerelor care nu le aparțin, copie date care nu sunt destinate publicității ș.a.m.d. Cu alte cuvinte, trebuie oare să distingem în continuare între meșteri și spărgători?

[Există și un substrat social al confuziei dintre meșter și spărgător. Marile firme au interesul să facă să dispară de pe piață meșterii care le-ar putea concura. Regretabil este însă când propaganda în favoarea marilor corporații pătrunde în manualele de etică și-și dă aere de doctrină filosofică, nu de activitate de *lobby*.]

Distincția dintre meșter și spărgător, în cazul tehnologiei informației, nu este ușor de trasat și pentru că totul depinde de posibilitatea de a cârpi, de a meșteri la produsele cumpărate. Dacă luăm de la magazin o încuietoare, în lipsa unei permisiuni din partea producătorului s-ar putea să nu avem voie să fabricăm o încuietoare de același tip. Nu ne trebuie însă permisiunea producătorului pentru ca să desfacem încuietoarea. Dacă se rupe cheia în încuietoare, n-avem încotro și spargem încuietoarea. Asta nu ne transformă însă în spărgători.

Se spune adesea că programele de computer diferă de încuietori. Programele de computer sunt texte. Seamănă cu cărțile dintr-o bibliotecă. Problema este că, spre deosebire de cărțile din bibliotecă, programele sunt citite nu de către noi, ci de către computer. Ca utilizatoare/utilizatori beneficiem doar de serviciile pe care ni le aduc aceste programe.

Atâta timp cât suntem utilizatoare/utilizator al programului doar computerul îl citește. Dacă ar fi să meșterim la program, să modificăm, să reparăm, să cârpim, atunci ar trebui să citim noi programul. Or, exact aici este nodul problemei.

Și încuietorile, și programele de computer ne fac niște servicii. Încuietoarea nu este nici ea importantă ca obiect fizic. Semnificativ este serviciul pe care producătorul l-a făcut posibil cu ajutorul încuietorii. La fel, programul, la care din pricina naturii sale nici nu ne uităm ca la un obiect fizic, face posibile servicii prin folosirea sa.

Meșterul, prin reparațiile sale sau modificările aduse produsului, nu face decât să adauge noi servicii sau să le facă din nou posibile pe cele vechi. Spărgătorul, în schimb, ne privează de toate serviciile sau de o parte dintre ele.

Unde apare atunci dificultatea trasării distincției dintre meșter și spărgător? În cazul unui pantalon pe care-l modificăm intervenim asu-

7. Etica și managementul informației

pra produsului, pe care-l scurtăm sau îl cârpim etc. Dar multiplicarea modificării respective nu se poate face lesne. Ne-ar trebui o fabrică pentru așa ceva. E limpede că atunci n-am mai fi meșteri. Cu totul altfel stau lucrurile în cazul unui program.

Dacă meșterim la un pantalon, putem comunica altora modificările făcute. Producătorul nu este adus însă într-o situație mai rea decât înainte pentru că oricine vrea să modifice pantalonul trebuie întâi să-l cumpere. Dacă meșterim la un program, este suficient să spunem cum se face acest lucru pentru ca și alții poată „cârpî“ în același mod programul. Atâta timp însă cât nu se pun în circulație alte exemplare decât cele realizate de producător situația este ca și-n cazul pantalonilor. Toată lumea învață să scurteze pantaloni, dar asta nu dă peste cap producția.

Etic vorbind, acțiunea de a meșteri și cea de a multiplica fără acordul producătorului programul sunt cât se poate de diferite. E limpede că, prin copierea de către alții a produsului, producătorul pierde posibilitatea de a vinde exemplarele realizate de el. Dar prin meșteritul la un program această posibilitate nu dispăre. Iar, atâta timp cât meșterul nu pune în circulație exemplare modificate, ci doar face cunoscute îmbunătățirile aduse, producătorul nu este adus într-o situație mai proastă decât era înainte. Tehnologic însă, în cazul programelor, spre deosebire de cel al pantalonilor, multiplicarea de exemplare de către alții nu poate fi blocată decât dacă este blocată și posibilitatea de a modifica programul. Practic asta înseamnă împiedicarea, printr-un mijloc sau altul, a citirii programului de către oameni.

Faptul că o persoană poate multiplica programe și fără să le citească, slujindu-se doar de programe de copiere este irelevant aici. Ceea ce este relevant e că meșterul *trebuie* să poată citi programul. Lectura programului deschide însă și posibilitatea de a crea alte exemplare din program. În schimb, dacă avem posibilitatea de a scurta un pantalon sau de a coase un nasture rupt prin aceasta nu ni se creează nici o posibilitate de a crea alți pantaloni și, cu atât mai puțin, de a-i pune în circulație.

De asemenea, faptul că în cazul cărților și oamenii de bună credință, și plagiatorii au acces la text este irelevant. În cazul cărții nu există *decât* lectura făcută de oameni. Fără ea cărțile ar fi obiecte ținute închise în rafturi, ceea ce ne-ar lipsi de principalul *serviciu* pe care ni-l pot aduce. În schimb, în cazul programelor de computer, deși acestea sunt texte, serviciul este oferit și atunci când doar computerul este cel care citește.

7.2 Varietatea licențelor

În cazul multor programe de calculator textul dat spre lectură calculatorului este altul decât cel scris, citit și modificat de către om. Pentru a se trece de la textul lizibil de către om la cel citit de mașină este nevoie de un program de traducere. Textul lizibil de către om se numește „sursă“.

Meșteritul la programele de computer este blocat foarte simplu dacă de vândut este vândut doar textul care poate fi citit de către mașină. Asemenea programe se numesc „cu sursă închisă“.

Atunci când este livrată doar sursa sau sursa dimpreună cu textul destinat mașinii se spune că avem de a face cu un program „cu sursă deschisă“.

Producătorii care vor să câștige de pe urma vânzării de programe optează pentru programele cu sursă închisă. Motivul este evident: dacă dai sursa, atunci singura protecție este brațul legii. Dacă ascunzi sursa, poți să împiedici copierea. Chiar dacă unii utilizatori sparg parolele și copie programul copie doar ceea ce poate citi numai mașina. Rămâne un secret modul în care arată sursa, iar asta înseamnă că alți producători nu pot ei să se inspire din surse. Or, lucrul acesta este cu mult mai important. Furtișagurile concurenței ar fi cu mult mai periculoase decât spargerile operate la nivelul utilizării programului.

Secretul comercial, nu copyright-ul sau patentul, este adevărata armă a producătorului care vrea să câștige din programe de calculator. Licențele sub care sunt puse aceste programe interzic de cele mai multe ori în mod explicit încercarea de a reconstitui sursele de la textul destinat mașinii.

S-ar părea că nu există alte surse de câștig decât vânzarea programului. Nu este deloc așa. Se poate câștiga în urma serviciilor pe care le oferi celor care folosesc programul. Acestea pot consta în ajutor la instalare, diverse adaptări ș.a.m.d. În aceste condiții, departe de a fi o utopie, există numeroase programe cu sursă deschisă.

Există și alte motive pentru a oferi programe cu sursă deschisă. Pentru unii faima pe care o dobândesc este mai importantă. Or, sursele sunt suprema dovadă a calității intelectuale a programului. Nu trebuie, de asemenea, subestimată dorința unora dintre cei care scriu programe de calculator de a face un cadou.

Licențele, în cazul programelor cu sursă deschisă, sunt de tipuri variate. Unele plasează programul pur și simplu în domeniul public. Majoritatea precizează însă condițiile în care poate fi folosită sursa. Este, de pildă, foarte important să știi ce statut trebuie să aibă modificările aduse sursei. Diversitatea este foarte mare. Practic singurul punct co-

7. Etica și managementul informației

mun este obligativitatea de a recunoaște calitatea de autoare/autor a persoanei care a scris sursa.

Chiar dacă n-o spun explicit, cei care acordă conotații negative termenului de meșter (*hacker*) sunt favorabili programelor cu sursă închisă. Altfel n-ai cum să înțelegi logica respingerii oricărui fel de *hacking*.

Există printre adepții surselor deschise și susținători ai cererii ca toate programele să fie cu sursă deschisă. Nici poziția aceasta nu rezistă. Ar însemna ca oricine scrie un program să fie obligat să divulge sursa, dacă vrea să pună în circulație textul destinat citirii doar de către computer. În acest fel ai forța însă oamenii să presteze un serviciu, acela de a oferi sursa. Ar fi relativ simplu de arătat că se încalcă în acest fel libertatea lor. Poziția obligativității de a face publice sursele devine astfel extrem de greu de susținut.

Alegerea licenței este o chestiune de contract între părți. Dacă vânzătorul oferă programul fără surse și cumpărătorul este de acord în mod voluntar, atunci lipsa accesului la sursă este legitimă. Este o prevedere a unui contract încheiat în mod liber de către părți.

La fel de important este ca, atunci când cineva folosește programe cu sursă deschisă, să respecte clauzele contractuale. Din nou este vorba despre un contract între părți.

Adevărata problemă apare însă atunci când sursa închisă asigură o poziție privilegiată pe piață. Un sistem de operare cu sursele închise face mai dificilă producerea de programe, care rulează sub sistemul respectiv, de către cei care nu cunosc secretele surselor sistemului. Etic însă acesta nu este argument suficient pentru a susține deschiderea surselor. Ar însemna să obligăm pe cineva să presteze în mod forțat un serviciu.

Poziția dominantă pe piață sau chiar poziția de monopol sunt condamnabile etic doar atunci când sunt obținute prin împiedicarea cu forța a intrării concurenților pe piață. Interzicerea accesului la sursele programului nu este însă o acțiune de forță. Condamnabilă cu adevărat, din perspectivă etică, este coluziunea cu statul în vederea obținerii poziției de monopol. Pot fi folosite diverse metode, inclusiv coruperea autorităților statului.

7.3 Implicații etice pentru managementul informației

Ce trebuie să facă în situația descrisă mai sus managerul unei firme? Ce fel de programe ar fi etic să folosească? Răspunsul de principiu este simplu: trebuie să respecte contractul încheiat.

Nu mai este așa de simplu de răspuns uneori dacă *transparența* prelucrării informațiilor devine o chestiune extrem de importantă. De pildă, programele care calculează rezultatul alegerilor ar trebui să fie cu sursă închisă sau cu sursă deschisă? Sursa închisă va lăsa întotdeauna loc la suspiciuni. Un program cu sursă deschisă este aici soluția etic mai bună pentru că asigură o deplină transparență.

Sunt multe situații în care achiziționarea sistemului de prelucrare a informațiilor se face cu bani publici. Ce soluție ar trebui să primeze atunci? Argumentul în favoarea surselor deschise este evident că, pe bani publici, ar trebui procurat un sistem care asigură prin licența sub care se află că orice beneficii, inclusiv cele ce țin de crearea unor noi programe, se vor întoarce acolo de unde provin banii, adică la cei care plătesc impozite. Este de văzut însă dacă se și găsesc programe cu surse deschise suficient de fiabile. Alegerea este departe de a fi simplă.

Mult mai dramatică este însă chestiunea accesului deschis sau închis la bazele de date. În vreme ce, în cazul unui programelor de prelucrare a datelor, argumentul în favoarea surselor deschise ține până la urmă de beneficiile legate de dezvoltarea cunoașterii, în cazul datelor care sunt supuse prelucrării raporturile par inversate.

Într-o etică în care persoana umană contează în primul rând ceea ce trebuie luat cu prioritate în considerare este răul pe care datele colectate pot să-l facă indivizilor.

O banală colecție de adrese de e-mail poate să fie cât se poate de periculoasă. Adresele pot fi folosite de cei care trimit *spam* și inundă căsuțele poștale electronice.

Alte principii decât cel al libertății individuale sunt adesea invocate. De pildă, în condițiile unor atacuri armate, fie ele și ale unor grupări mici, este invocat un principiu al securității. Acest principiu poate duce la drastice limitări ale libertății individuale.

În practică, managerii trebuie de multe ori să găsească o cale pentru a echilibra cumva respectul pentru principii incompatibile precum cele amintite mai sus.

Folosind exercițiile și analizele de caz din secțiunea finală puteți găsi numeroase materiale despre dilemele etice în fața cărora sunt puse

7. Etica și managementul informației

persoanele care iau decizii cu privire la strângerea, prelucrarea și accesul la informații.

7.4 Etica și managementul informației: termeni-cheie

- informații personale
- *hacker*
- *cracker*
- program cu sursă închisă
- program cu sursă deschisă
- licențe pentru programe de calculator
- etică centrată pe libertatea individuală
- etică centrată pe securitatea colectivității
- acces la baze de date

7.5 Etica și managementul informației: exerciții și analize de caz

Exercițiul 1: Etica este un domeniu al filosofiei cu incredibil de multe ramificații. Accesați situl *Stanford Encyclopedia of Philosophy* <<http://plato.stanford.edu/>>. Introduceți în caseta „Search the SEP“ expresia `title:"ethics"` pentru a limita căutarea cuvântului „*ethics*“ la titlurile articolelor. Încercați să vă faceți o idee despre variatele tipuri de etică. Citiți articolul „Computer Ethics“ de Terrell Bynum pentru a vă face o idee despre conceptele de bază din etica tehnologiei informației.

Analiza de caz 1: Margo Anderson și William Seltzer folosesc expresia „*the dark side of numbers*“ care ar putea foarte bine să fie titlul unei serii de analize de caz. Porniți de la Margo Anderson și William Seltzer „Official Statistics and Statistical Confidentiality: Recent Writings and Essential Documents“ <<http://www.uwm.edu/~margo/govstat/integrity.htm>> pentru a găsi surse pentru o analiză de caz.

7. Etica și managementul informației

De pildă, puteți folosi articolul lui William Seltzer, "Population Statistics, the Holocaust, and the Nuremberg Trials," *Population and Development Review*, vol. 24, nr. 3. (septembrie 1998), pp. 511-552 <<http://www.uwm.edu/~margo/govstat/seltzer.pdf>> pentru a analiza cazul folosirii statisticilor demografice de către naziști în timpul Holocaustului.

Exercițiul 2: Clarificați sensurile opuse ale termenului *hacker*. Porniți de la articolul din Wikipedia, „Hacker“ <<http://en.wikipedia.org/wiki/Hacker>>.

Exercițiul 3: Porniți de la articolul din Wikipedia despre *crackers* <http://en.wikipedia.org/wiki/Black_hat> pentru a clarifica istoria termenului *cracker* (*Black Hat*).

Analiza de caz 2: Editura O'Reilly are o serie întreagă de cărți intitulată „O'Reilly Hacks“. Găsiți situl acestei serii de cărți la adresa <<http://www.oreilly.com/hacks/>>. Pe sit, prezentarea fiecărei cărți include extrase din text. Alegeți unul sau mai multe dintre aceste *hacks* și construiți analize de caz. Încercați să vedeți dacă este sau nu rezonabilă din punct de vedere etic acțiunea descrisă în *hack*. De unde - și pentru cine - ar putea veni vreun pericol?

Exercițiul 4: Analizați tipuri de licențe care sunt incluse în contractele dintre producătorii de programe de calculator și beneficiare/beneficiari. Porniți de la articolul „Software license agreement“ din Wikipedia <http://en.wikipedia.org/wiki/Software_license_agreement>. Acordați atenție disputei cu privire la neutralitatea articolului.

Analiza de caz 3: Analizați cazul „ProCD v. Zeidenberg“ sau un caz similar. Pentru faptele din cazul „ProCD v. Zeidenberg“ porniți de la Wikipedia <http://en.wikipedia.org/wiki/ProCD_v._Zeidenberg>.

Exercițiul 5: Analizați conținutul unora dintre licențele pentru programele cu sursă deschisă. Porniți de la categoria de articole „Open source licenses“ din Wikipedia <http://en.wikipedia.org/wiki/Category:Open_source_licenses>. Citiți și conținutul licențelor, nu doar articolul de enciclopedie.

Analiza de caz 4: Pornind de la Wikipedia <<http://en.wikipedia.org/wiki/Gpl-violations.org>> analizați cazul *Fortinet*. Găsiți eventual alte cazuri de folosire ilegală și/sau non-etică a programelor cu sursă deschisă.

Analiza de caz 5: Articolul din Wikipedia despre firma Fortinet susține că produsul *FortiGuard* ar fi folosit de către regimul din Myanmar (Birmania) pentru a bloca siturile Internet care critică dictatura militară din țara respectivă <<http://en.wikipedia.org/wiki/Fortinet#Controversies>>. Și alte firme din domeniul computerelor au fost criticate pentru serviciile făcute unor regimuri opresive. Este

7. Etica și managementul informației

vreo problemă de ordin etic dacă o firmă care se ocupă cu tehnologia informațiilor vinde produse care pot fi folosite de către regimuri dictatoriale în scopuri represive?

Analiza de caz 6: Tipul de mijloace folosite de *cracker*-i poate fi utilizat și de către poliție, de pildă pentru a face o percheziție la distanță (în computere legate la Internet). Exemple de astfel de programe găsiți în Wikipedia <<http://en.wikipedia.org/wiki/Policeware>>. În Germania există o dezbatere cu privire al perchezițiile *online* <<http://de.wikipedia.org/wiki/Online-Durchsuchung>>. Analizați cazul unui asemenea mijloc de percheziție. Există motive pentru a avea dubii de ordin etic cu privire la asemenea mijloace de a percheziționa computere?

Analiza de caz 7: Atunci când folosiți un motor de căutare pe Internet este tehnic posibil să se stocheze date cu privire la ceea ce ați căutat. Asemenea dat pot fi vândute apoi, de pildă, unor firme de marketing. Uniunea Europeană a criticat Google pentru modul în care păstrează date despre cei care caută informații pe Internet (a se vedea <http://en.wikipedia.org/wiki/Criticism_of_Google>; trimiterile erau acolo când le-am verificat pe date de 2007-09-30). Analizați acest caz sau unul similar.

Bibliografie

- [1] Peter Drucker. *Management: Tasks, Responsibilities, Practices*. New York: Harper & Row, 1985. Citată la pp. 11 și 27.
- [2] Peter Drucker. Managing knowledge means managing oneself. *Leader to Leader*, nr.16, 2000. <<http://www.leadertoleader.org/knowledgecenter/L2L/spring2000/drucker.html>>. Citată la p. 13.
- [3] Ben Forta. *SQL pentru începători*. București: Teora, 2002. Citată la pp. 70 și 71.
- [4] D. M. Griffiths. *Managing Information - A Practical Guide*. <<http://www.managing-information.org.uk/introduction.htm>>, 2006-03-15. Citată la pp. 6, 7 și 8.
- [5] Rob Mattison. *Understanding Database Management Systems*, ediția a 2-a. New York: McGraw-Hill, 1997. Citată la p. 29.
- [6] A. I. Mihailov, A. I. Ceornîi și R. S. Ghilearevski. *Informatica documentară*. București: Editura Științifică, 1970. Citată la p. 6.
- [7] Chris Newman. *SQLite*. Indianapolis: Sams, 2005. Citată la p. 71.
- [8] Ioan Radu, Minodora Ursăcescu și Florin Ioniță. *Informatică pentru managementul firmei*. București: Editura „Tribuna Economică”, 2001. Citată la p. 69.
- [9] Meenakshi Raman și Sangeeta Sharma. *Technical Communication: Principles and Practice*. New Delhi: Oxford University Press, 2004. Citată la p. 87.
- [10] Herbert Schnädelbach. *Introducere în teoria cunoașterii* (traducere de Mihail Radu Solcan). Pitești: Paralela 45, 2007. Citată la p. 22.
- [11] Mihail Radu Solcan. *Eseul filosofic*. București: Editura Universității din București, 2004. <<http://www.ub-filosofie.ro/~solcan/eft/index.html>>. Citată la pp. 20, 25, 89, 90 și 96.

BIBLIOGRAFIE

- [12] George Soros. *The Alchemy of Finance*. New York: Wiley, 1994.
Citată la p. 44.