
1. Introducere în Informatica Medicală

1.1. *Obiectul Informaticii Medicale*

1.1.1. *Scurt istoric*

A. Informatica medicală este o disciplină tânără, termenul apărând în cursul anilor '60. În accepțiunea inițială informatica medicală cuprindea *programele de calculator* cu aplicabilitate în domeniul medical. Progresul tehnic rapid a arătat însă că, pentru aceleași aplicații, atât programele cât și suportul fizic se schimbau; ceea ce rămânea la fel era modul în care era prelucrată *informația*.

B. Astfel, în accepțiunea actuală, centrul definiției s-a mutat de la „calculator” la informație. Coiera [1997] chiar atrage atenția în acest sens: „Informatica medicală se ocupă de calculatoare tot atât de mult cât se ocupă cardiologia de stetoscoape”.

1.1.2. *Obiectul informaticii medicale*

Accepțiunea clasică: totalitatea programelor de **calculator** cu aplicații în domeniul biomedical și sănătate.

Definiția actuală: disciplina care studiază întregul flux al **informației medicale**: generare, achiziție, stocare, transmitere, prelucrare și utilizare.

1.2. *Teoria informației*

1.2.1. *Noțiunea de informație*

Pentru a ne ocupa de *informația medicală*, să încercăm mai întâi să privim conceptul de *informație* la modul general.

A. Termenul de informație este folosit în mod curent în viața de zi cu zi, fiind cel mai adesea asociat cu aducerea unui element de noutate. Fiind un concept cu grad înalt de generalitate (categorie filosofică), informația nu poate fi definită în manieră clasică, pornind de la genul proxim și precizând diferențele specifice, ci prin proprietatea sa esențială – cea de a înlătura o nedeterminare.

B. Noțiunea de informație

Informația este un concept cu grad înalt de generalitate caracterizat prin proprietatea de a înlătura o nedeterminare (incertitudine).

1.2.2. *Proprietățile informației*

Informația nu este materie; totuși ea nu poate exista înafara materiei. Norbert Wiener spunea “Creierul nu secretă informație precum ficatul fiere”.

Informația nu este energie; totuși ea nu se poate transmite fără un suport energetic.

Nu este o relație directă între cantitatea de energie ce însoțește transmiterea unei informații și cantitatea de informație transmisă. De exemplu, energia unui trăsnet în timpul unei furtuni este imensă, însă informația transmisă este neglijabilă; în schimb un foșnet într-o pădure, purtat de o energie infimă poate reprezenta o informație vitală – punând pe fugă un animal! Să semnalăm totuși că nu există relație nici între cantitatea de informație și efectele sale. De ex. Legenda lui Tezeu din metodologia greacă: Tezeu promisese tatălui său Aegeus, că dacă va învinge în luptă minotaurul va înlocui pânza neagră a corabiei cu pânză albă, dar a uitat și tatăl său s-a aruncat de pe stânci. Informația primită a fost doar 1 bit.

1.2.3. Triada abordărilor complete

A. Introducerea aspectelor informaționale în studiul materiei vii completează imaginea noastră privind complexitatea sistemelor biologice, actualmente considerându-se că o abordare completă trebuie să acopere atât aspectele materiale și energetice cât și cele informaționale.

B. Triada abordărilor complete

- aspectul material – structura
- aspectul energetic – suportul funcțional
- aspectul informațional – mecanismul funcțional

C. Valoarea utilă a informației depinde de receptor.

Sensul noțiunii de informație, așa cum a fost prezentat mai sus etc. legat de altă noțiune – nedeterminarea (sau incertitudinea) – vag definită la rândul său. Același mesaj poate să aibă valori informaționale diferite pentru diferiți receptori: pentru o persoană care deja știa conținutul său cantitatea de informație primită este zero, însă pentru receptorii care nu-i știau conținutul va putea fi evaluată cantitatea de informație primită.

1.2.4. Cantitatea de informație

Parcurgerea acestui subiect necesită cunoștințe fundamentale de teoria probabilităților

A. Pornind de la proprietatea fundamentală a informației, cea de a înlătura o nedeterminare, Shannon a considerat că informația primită este invers proporțională cu probabilitatea de apariție a evenimentului: dacă se va întâmpla un eveniment cu probabilitate mare, informația primită este mică; în schimb primim o informație “mai mare” dacă apare un eveniment mai rar. (Ziariștii exploatează intens această relație!)

B. Relația propusă de Shannon pentru calculul cantității de informație care este primită când se petrece un eveniment cu probabilitatea p_i cuprinde logaritmul în baza 2 din inversul probabilității p_i .

- Cantitatea de informație eliberată de un eveniment a cărui probabilitate este p_i

$$I_i = -\log_2 p_i \quad (1.2.4.a)$$

C. Pe baza acestei relații stabilește unitatea de măsură pentru cantitatea de informație, numită *bit* (de la Binary digiT).

Definiție: Un bit este cantitatea de informație primită când se înlătură o nedeterminare de $\frac{1}{2}$

D. În mod uzual informația se transmite printr-o succesiune de evenimente, numită adesea *mesaj*, iar un eveniment într-un mesaj se mai numește *simbol*.

În cazul unui mesaj format din N evenimente, fiecare eveniment i apare de n_i ori, aducând de fiecare dată informația I_i , deci mesajul aduce informația.

$$I = n_1 I_1 + n_2 I_2 + \dots + n_k I_k = \sum_{i=1}^k n_i I_i \quad (1.2.4.b)$$

E. Valoarea medie a informației corespunzătoare unui eveniment într-un șir de N evenimente se mai numește “entropie informațională, H , și se calculează astfel:

$$H = \frac{n_1 I_1 + n_2 I_2 + \dots + n_k I_k}{N} = \frac{n_1}{N} I_1 + \frac{n_2}{N} I_2 + \dots + \frac{n_k}{N} I_k$$

la limită (i.e. atunci când $N \rightarrow \infty$) relația devine:

$$H = p_1 I_1 + \dots + p_k I_k = \sum_{i=1}^k p_i I_i$$

Înlocuind I_i conform relației (1.2.4.a), obținem formula (1.2.4.c), formulă fundamentală în teoria informației, numită și formula lui Shannon pentru entropia informațională.

- Entropia informațională este cantitatea medie de informație per eveniment (simbol) într-un mesaj:

$$H = - \sum_{i=1}^k p_i \log_2 p_i \quad (1.2.4.c)$$

1.2.5. Relația între entropia informațională și entropia termodinamică

A. Termenul de “entropie” a fost introdus în termodinamică pentru enunțarea principiului al II-lea al termodinamicii: “În procesele termodinamice entropia nu poate să scadă: ea rămâne constantă în cadrul proceselor reversibile și crește în cazul proceselor ireversibile”.

B. Relația între entropia termodinamică și cea informațională poate fi înțeleasă pornind de la experimentul “ideal” propus de Maxwell pentru explicarea variației entropiei în cazul proceselor ireversibile, prezentat în cadrul cursului de biofizică.

C. Se vede deci că sistemul poate evolua în sens contrar celui dictat de al II-lea principiu al termodinamicii în cazul în care primește o informație. Acesta este mecanismul prin care sistemele vii evoluează spre stări tot mai organizate și deosebite de mediul înconjurător.

1.2.6. Redondanță

A. Entropia informațională are valoare maximă când evenimentele din mesaj sunt echiprobabile: $p_i = \frac{1}{k}$;

atunci $H_{\max} = k \left(\frac{1}{k} \log \frac{1}{k} \right)$ de unde se obține relația (1.2.6.a)

$$H_{\max} = \log_2 k \quad (1.2.6.a)$$

B. Un exemplu ar fi cazul unui mesaj encriptat, în care probabilitatea apariției unui simbol este (cel puțin aparent) independentă de simbolurile anterioare. În mesajele reale probabilitatea unui simbol depinde de simbolurile anterioare; putem, în funcție de context, să “ghicim” ce urmează, putem folosi prescurtări, putem observa greșeli cum ar fi omisiunea unei litere etc. Deci informația nu este distribuită uniform în mesaj sau chiar în interiorul cuvintelor, cantitatea de informație transportată în realitate fiind inferioară celei maxime ce ar putea fi transmise folosind aceeași lungime a textului. Această diferență, între cantitatea maximă ce poate fi conținută în mesaj și cea reală se numește redondanță și reprezintă o parte din mesaj care... nu conține informație!

C. Relația de definiție a redondanței absolute este:

$$R = H_{\max} - H_{\text{real}} \quad (1.2.6.b)$$

Raportând redondanța absolută la $H_{\max}$ se definește redondanța relativă.

$$R_r = R / H_{\max} \quad (1.2.6.c)$$

D. *Rolul redondanței*

Aparent redondanța ar reprezenta o încărcătură inutilă în mesaj. Totuși, prezența ei diminuează rolul negativ al perturbațiilor ce apar în cursul transmiterii informației, putând deseori reconstitui mesajul inițial chiar dacă unele simboluri au fost perturbate.

1.3. Transmiterea informației

1.3.1. Sisteme de comunicație

A. Am precizat anterior că valoarea utilă a informației depinde de receptor, deci noțiunea de informație are sens doar dacă se transmite; altfel, rămâne în faza de “informație potențială”.

B. Transmiterea informației presupune o *sursă* a informației (emițător E) și un destinatar (receptor R). Spațiul dintre S și R reprezintă *canalul de comunicație* (C). Pe canalul de comunicație pot să apară diverse *zgomote* care perturbă sistemul de comunicație afectând calitatea transmisiei (figura 1.3.1.a).

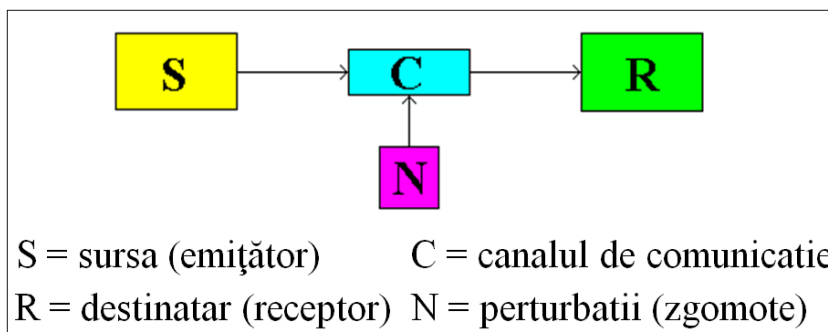


Fig. 1.3.1.a. Sistem de comunicație simplu

C. Să introducem doi termeni importanți în cadrul sistemelor de comunicație:

- *mesaj* – un termen pe care îl folosim când ne referim la conținutul informațional al transmisiei
- *semnal* – suportul fizic care transportă mesajul (sunet, curent electric etc.).

D. Pentru diminuarea efectelor perturbațiilor sau pentru a asigura transmiterea mesajului la distanțe foarte mari, se introduc pe canalul de transmisie niște dispozitive numite *traductori*. Un traductor schimbă suportul fizic al unui semnal (figura 1.3.1.b). De exemplu, în cazul unei convorbiri telefonice, microfonul este traductorul localizat lângă emițător, transformând sunetele (variații ale presiunii aerului) în variații ale unui curent electric. Canalul de comunicație este reprezentat de firele telefonice. La destinație un alt traductor, casca telefonică, transformă variațiile curentului electric în vibrații ale unei membrane elastice generând astfel sunete.

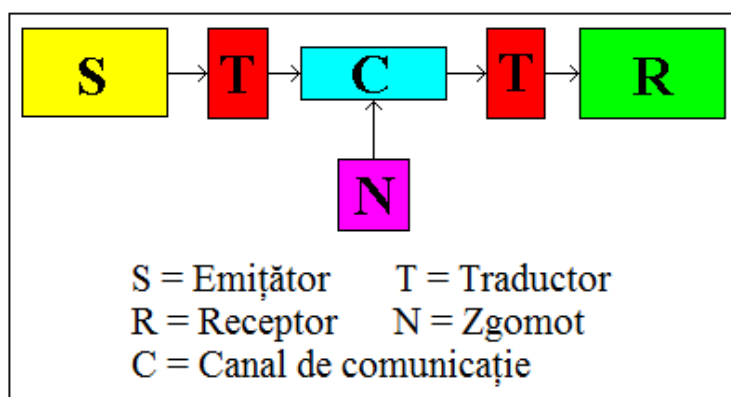


Fig. 1.3.1.b. Schema generală a unui sistem de comunicație

E. Există și alte dispozitive ce pot fi utilizate în sisteme de comunicație, de ex. *modem-ul*. Denumirea modem provine de la modulator/demolator. Modemul este un dispozitiv care asigură modularea semnalului, adică suprapunerea semnalului real peste un semnal purtător (undă purtătoare) care are caracteristici încât se diminuează efectul perturbațiilor (de ex. perturbațiile uzuale, de joasă frecvență, sunt eliminate dacă unda purtătoare are frecvență înaltă).

F. O altă transformare pe care o putem aplica semnalului pentru transmisie este *codificarea*. Mesajul este compus uzual dintr-o succesiune de simboluri. Totalitatea simbolurilor utilizate pentru a compune un mesaj se numește *alfabet*. Simbolurile alfabetului se mai numesc “*litere*”, iar cu literele putem construi *cuvinte*. Totalitatea cuvintelor cu sens reprezintă un *dicționar*, iar precizarea sensului cuvintelor se numește *semantică*. Cu ajutorul cuvintelor se pot construi propoziții; regulile de construcție a propozițiilor se numește *sintaxă*. Un dicționar împreună cu semantica și o sinteză reprezintă un *limbaj*. Noi folosim uzual pentru comunicație *limbaje naturale*, dar există posibilitatea utilizării unor *limbaje formale* sau *artificiale*. Diferitele componente ale sistemului de comunicație pot folosi diferite alfabete sau dicționare. Transpunerea unui mesaj dintr-o formă ce utilizează un alfabet într-o formă în alt alfabet, cu anumite reguli de corespondență se numește *codificare*. Operațiunea inversă se numește *decodificare*. Transpunerea unui mesaj dintr-un limbaj în altul se numește *traducere*.

G. Să mai menționăm legat de sistemele de comunicație că există o capacitate limitată de transmisie a informației pe canalul de comunicație, numită *viteză de transmisie*, măsurată în bit/secundă.

1.3.2. Exemple de transmisie a informației în materie vie

A. *Codul genetic.* Informația privind structura proteinelor ce pot fi sintetizate este stocată în molecula de AND din nucleu. Acizii nucleici conțin 4 baze azotate: adenina A, timina T, citozina C și guanina G (în cazul ARN în loc de timină apare uracilul u). Proteinele sunt formate din 20 de aminoacizi esențiali. O succesiune de 3 baze azotate din AND se numește *codon* și poartă informația pentru codificarea unui aminoacid într-o secvență proteică. Totalitatea corespondențelor între codoni și aminoacizii corespunzători poartă denumirea de *cod genetic*. Porțiunea dintr-un lanț AND care poartă informația pentru sinteza unei proteine se numește *genă*, iar ansamblul tuturor genelor unei specii se numește genom. Genomul uman conține circa 30.000 gene.

B. *Codificarea informației în sistemul nervos.* Pe axoni informația este transmisă printr-o succesiune de impulsuri nervoase; fiecare impuls nervos este un potențial de acțiune care are întotdeauna aceeași amplitudine. Unui stimul mai intens îi corespunde o rată mai ridicată de formare a potențialelor de acțiune; spunem că pe axon informația privind intensitatea stimulului este *codificată în frecvență*.

La nivelul sinapselor are loc o descărcare a veziculelor cu mediator chimic în spațiul sinaptic, cantitatea de mediator descărcată fiind proporțională cu frecvența impulsurilor nervoase pe axon; spunem că în spațiul sinaptic informația privind intensitatea stimulului este *codificată în amplitudine*, aceasta fiind reprezentată de cantitatea de mediator descărcată.

La nivelul membranei postsinaptice, mediatorul se cuplează pe receptorii postsinaptici, se deschid canalele de sodium, membrana se depolarizează și apare un potențial care se propagă pe membrana corpului neuronal sau pe dendrite. Spunem că informația este *codificată în amplitudine*, aceasta fiind reprezentată de potențialul local.

1.4. Informatica medicală

După această incursiune în teoria informației putem reveni la noțiunea centrală din informatică medicală și anume *informația medicală*.

Ce este informația medicală și când apare ea?

1.4.1. Date și cunoștințe

Să încercăm să schițăm în cel mai simplificat mod actul medical primar și anume vizita pacientului la medic. Poziția centrală în activitate medicală este ocupată de *pacient*. Fără pacient nu există medicină! Actorul principal al activității medicale este *medicul*, dar în activitatea medicală sunt implicate numeroase alte persoane care aparțin așa-numitelor "*profesii aliate*". Dialogul medic-pacient începe uzual cu expunerea de către pacient a motivelor pentru care s-a prezentat la medic. Această descriere reprezintă transmiterea unor informații de la pacient către medic. Informațiile care se transmit sau se utilizează într-un act medical (sau ca urmare a unui act medical) reprezintă *informația medicală*. Dialogul este succedat de către examenul obiectiv al pacientului, medicul colectând astfel și alte informații despre pacient. Să observăm că

aceste informații au un caracter individual – sunt valabile pentru *acest pacient*. Aceste informații se numesc **date**. Uzual paleta datelor se completează cu informații provenind și din alte investigații (probe de laborator, explorări funcționale, radiografii etc.). Indiferent cât de complexe ar fi ca reprezentare, ele sunt “date”, fiind caracteristice unui anumit individ.

În același timp, medicul utilizează și alt fel de informații, numite *cunoștințe*. Acestea au un caracter general și sunt acumulate în cursul pregătirii profesionale precum și în experiența sa practică. Fără aceste cunoștințe informațiile sub formă de date nu pot fi interpretate (revenim la afirmația că valoarea utilă a informației depinde de receptor; practic, fără aceste cunoștințe receptorul datelor nu este “medic”). De aceea numeroși autori numesc *informație* doar *datele interpretate*. Pentru a evita confuzia între termenul *informație* folosit la modul general și *informație* pentru treapta de *date interpretate*, vom păstra termenul de *date interpretate* pentru acest nivel.

1.4.2. Ciclul elementar al informației medicale

Prin interpretarea datelor de către medic pe baza cunoștințelor sale, este generată o nouă informație numită diagnostic. Pe baza diagnosticului, folosind din nou cunoștințele sale, medicul stabilește un plan terapeutic pe care îl aplică pacientului cu scopul de a îmbunătăți starea pacientului. Urmărirea evoluției pacientului este însoțită de colectarea unor noi informații sub formă de date. Se observă că se încheie un ciclu al fluxului informațional în activitatea medicală, numit “ciclul elementar al informației medicale” (figura 1.4.2).

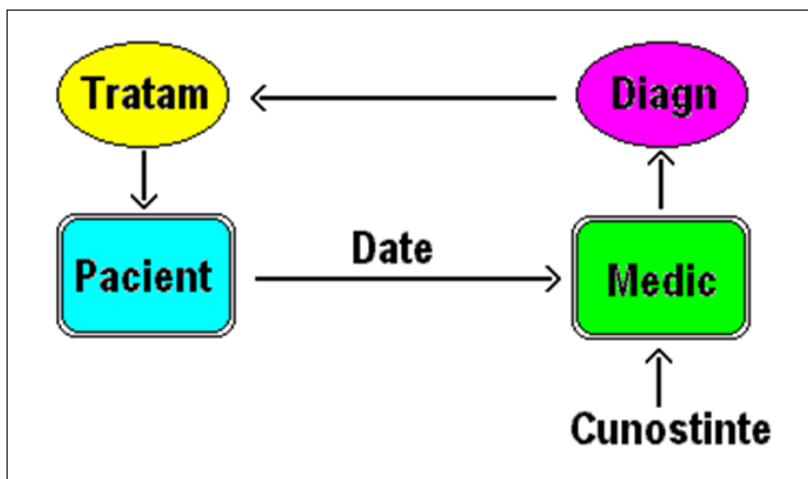


Fig. 1.4.2. Ciclul elementar al informației medicale

1.4.3. Tipuri de date

Informațiile culese despre starea pacientului, adică datele, pot îmbrăca diverse forme:

- *date calitative* – cu caracter descriptiv, așa cum apar în anamneză
- *date numerice* – forma uzuală de prezentare a rezultatelor de laborator
- *grafice* – modul de înregistrare a evoluției în timp a unor mărimi biologice (ex.: semnalul ECG, EEG etc.)

-
- *sunete* – de ex. fonocardiograma; modul de prelucrare este asemănător cu cel al altor semnale
 - *imagini* – radiografia, tomografia, ecografia etc.
 - *imagini dinamice* – filme.

Modul de achiziție, stocare și prelucrare este specific pentru fiecare tip de date și în cadrul cursului nostru le corespund capitole separate.

1.4.4. Tipuri de cunoștințe

Cunoștințele pot fi de mai multe feluri:

- *cunoștințe explicite* – care se pot formaliza, se pot exprima în propoziții, pot fi ușor transmise pe cale orală sau scrisă
- *abilități* sau *cunoștințe tacite* (limba engleza - *skill*) – cele câștigate prin experiență practică (de ex. îndemânarea unui chirurg sau a unui dentist); nu pot fi transmise ușor.

1.4.5. Operații cu informații

Urmărind ciclul de viață al informației, din momentul generării sale până în momentul utilizării, observăm că informația suferă o serie de operații:

- *achiziția (colectarea)* – presupune mijloace specifice tipului de informație
- *stocarea* – baze de date, respective baze de cunoștințe
- *transmitere* – căi, procedee
- *prelucrare* – cu o largă paletă de metode specifice, pentru a extrage elementele esențiale în vederea interpretării și utilizării
- *protecție* – măsurile ce se impun pentru asigurarea integrității informației stocate sau transmise, precum și a confidențialității acestora
- *interpretare/utilizare* – pasul final, în care informația este integrată în acțiunile specifice nivelului.

1.5. Capitolele informaticii medicale și structura cursului

1.5.1. Clasificarea informației medicale pe nivele structurale

În ciclul elementar al informației medicale prezentat mai sus am luat în considerare informațiile care apar în activitatea medicală curentă, la nivelul individului, numit pacient. Totuși fenomenele care se petrec în materia vie (legate de starea de sănătate a pacientului) privesc deseori nivele infraindividuale, pornind de la nivelul molecular sau celular, urcând prin nivelul de țesut, organ sau sistem până la nivelul întregului organism sau nivelul individual.

Pe de altă parte, activitatea medicală este organizată în unități care prestează servicii pentru populație, deci putem urmări fluxul informațional și la nivel supraindividual, de comunitate. Corespunzător acestor nivele structurale avem diferite discipline biomedicale precum și diferite capitole corespunzătoare ale informaticii medicale.

1.5.2. Bioinformatica – relații cu alte discipline

Bioinformatica este un capitol de sine stătător al informaticii medicale. Însă, prin specificul obiectului său de a urmări modul în care este reprezentată informația în structurile vii, obiectul său este strâns legat de o serie de discipline privind structura și funcția structurilor din materia vie. Astfel, sunt necesare cunoștințele de biofizică, biochimie, biologie celulară și moleculară precum și genetica.

Din punct de vedere al abaterilor formale, legătura cu teoria probabilităților, biostatistica, teoria sistemelor și chimia cuantică este evidentă. De asemenea, se folosesc diverse limbaje de programare – R, PERL, C++, Java etc.

După cum vedem, se poate afirma – pe bună dreptate – că bioinformatica este una dintre cele mai complexe discipline, pregătirea specialiștilor în acest domeniu fiind cel mai adesea inclusă în categoria “studii avansate”.