

CERCADORS

Carles UDINA i COBO

carles.udina@cat21.cat

2007-01-25. Rev. 2008-12-11

Es recomana al lector que vagi a "[Ajuts \(=Help's\) dels programes informàtics](#)" i llegeixi els primers paràgrafs, els de "L'ordre més comú i natural" i així no ho repetirem aquí.

Com també es reitera al final, el tema aquí plantejat és part d'un problema més global, el de la classificació i recuperació de la informació, que es tracta a "[Classificacio.pdf](#)". Aquest és doncs, un document complementari.

ANTECEDENTS DELS ACTUALS CERCADORS

La química és una de les ciències més recents, si més no comparativament amb altres ciències naturals iniciades molts més segles abans. La química només va poder-se iniciar a partir del coneixement dels nivells estructurals de la matèria, coneixement que va proporcionar la progressiva construcció de la "Taula periòdica dels elements químics" (els "Àtoms"), de MENDELEYEV (amb la "Llei periòdica", el 1869) i MOSELEY. Aquests nivells són:

- El nivell de les substàncies,
- el nivell inferior de les molècules que conformen substàncies i
- el nivell dels àtoms que conforma les molècules.

Anteriorment, conceptes no estructurats com "Aigua", "Aire", "Terra" i "Foc", o els apriorismes de l'Alquímia no permetien coneixements comprensius, només un coneixement descriptiu i amb equívocs.

Això és una qüestió aparentment diferent a la que aquí es tracta, però cal fer èmfasi perquè veurem finalment que també està íntimament relacionada, per exemple, els actuals problemes per a desenvolupar les Ciències del coneixement o el sonat fracàs, l'any 1992, de la mal anomenada "Intel·ligència artificial".

El desenvolupament de la química va ser espectacular en el segle XX per raó de les nombrosíssimes molècules, les seves propietats, aplicacions industrials i mèdiques i els corresponents processos de síntesis. Per això va ser en l'àmbit de la química on es van haver de desenvolupar els primers sistemes potents/eficients de recuperació de la informació.

Tot i els nombrosos "Thesaurus" en ús, apareguts a partir de finals del XIX i enfocats a definir continguts temàtics, les eines més eficaces i emprades pels químics van basar-se en la transitivitat de les cites/referències, uns aspectes extrínsecs al contingut informatiu. Els anys 70' el "Science Citation Index" i el "Permuterm Subject Index" eren l'eina bàsica de recerca d'estudiants i investigadors.

La progressiva implantació i millora de la informàtica –especialment els "Bancs de documentació científica" ("PASCAL", "ENVIROLINE", ...)– va reactivar l'ús dels "Thesaurus", i sobre tot l'aparició de nous "Thesaurus" amb estructures conceptuals més actualitzades però en qualsevol cas, marcadament subjectives.

A més a més, aquesta mateixa proliferació va crear el problema de les traduccions entre els diferents "Thesaurus" en ús –pràcticament, cada Banc de dades tenia el seu propi "Thesaurus"–, greu inconvenient si es pretenia la lògica consulta a més d'un Banc de dades.

Cal afegir, a més a més, que els "Thesaurus" seguien essent llistes fetes per documentalistes, i que un cop fetes s'entraven en l'ordinador, però no un conjunt d'elements interrelacionats i processables per l'ordinador segons uns criteris comprensius ("Metadades"*). Llevat desconeixement d'altres iniciatives similars, la primera iniciativa comprensiva i relacional d'aquest tipus la vaig desenvolupar el 1987** mitjançant un programa informàtic que simulava els endomorfismes (o els canvis de base) d'un espai algebraic, i construïa "Successions algebraiques exactes curtes", que generaven successions topològiques de CAUCHY, però no amb números, sinó que amb els conceptes.

* Els "Metadades" han estat, a partir dels 90', la base d'eines molt reeixides en l'àmbit empresarial com els DataWarehouse, avui englobats en la "Business Intelligence".

** DICSEM (DICcionari SEMàntic), que poc després va donar pas al nucli del "[Sistema conceptual \(intrínsec i exacta\)](#)".

Però els anys 90',

- d'una banda per les limitacions dels "Thesaurus", i
- per una altra banda per l'evolució i economia del "hardware" –grans velocitats i capacitats d'emmagatzemament–,

es van deixar de banda les tècniques semàntico-estructurals dels "Thesaurus" enfocades al contingut, i es va retornar a les tècniques basades en la citació/ referenciació transitiva, però amb el suport de la potentíssima gestió informàtica. La velocitat dels microprocessadors fins i tot han permès ressuscitar les recerques alfabètiques. Qualitativament, un pas al darrera. Tot això tan enumeratiu és la base del que avui anomenem "Cercadors", com els prou coneguts de "Google", "Yahoo", "MSN", ...

La referenciació/ citació de "Pàgines Web" ha portat finalment a iniciatives com la "Web semàntica" i els "Ontology Web Languages" (= "OWL"). Més solucions parcials que creen la necessitat de més solucions complementaries... No és clar si cada vegada hi ha més solucions, o més coses pendents de solucionar.

Els increments de la velocitat de processament, dels processos manuals de fa només 40 anys, als milions de Herz en els ordinadors actuals, ha permès aquest reeixit tractament enumeratiu, no comprensiu, de la informació. Però el problema ha reaparegut quan la quantitat d'informació que s'ha anat penjant d'aquets processos també ha augmentat exponencialment, doncs som mils de milions de persones potencialment aportadores d'informació.

NECESSITAT D'UNA NOVA TECNOLOGIA

El que és clar, és una percepció compartida: cal alguna nova tecnologia. Com deia un premi Nobel*, asincerant-se, respecta d'una altra situació similar –la mecànica quàntica– "ens cal una nova teoria que avui encara no ens podem ni imaginar".

* Antoni LEGGETT ([El País, 2005-06-18](#))

Una de les primeres barreres comença pels idiomes*. I fins i tot amb un mateix idioma per les diferents cultures. Un consultor de "SAP AG"*** exposava en un congrés d'aquesta tecnologia, que tot i parlar tothom l'anglès, el pitjor problema dels equips de treball –i que ni tan sols s'imaginava com es podia resoldre– no era tecnològic sinó que semàntic, eren els diferents "idiomes" que parlava tothom.

* Al traduir automàticament una pàgina Web, la primera sorpresa és la traducció de "Home" (= "Pàgina principal") per "domèstic". No es tracta de corregir aquest error, perquè després d'aquest apareix un altra de similar i després un altre, ..., i així fins a l'infinit. Es tracta de replantejar –una esmena a la totalitat– les actuals metodologies de traducció.

*** Kai KUTTENKEULES (ICM Software & Management Consulting AG). SAP Meeting (Barcelona, 2000)

Un exemple podria ser un àmbit qualsevol, l'aprenentatge a distància en centro-amèrica en aquest cas, on una expressió inofensiva com "donar d'alta en el campus" s'entén com "matar y enviar al camposanto". O entre "coger" i "joder" (Argentina), etc.

El lector hauria de perdre dos minuts en llegir les reflexions i preguntes inicials ("¿Què coneixem del coneixement? Estat actual i reflexions" [0]) de l'escrit "L'estructura del pensament i del coneixement". Possiblement això li ajudi a entendre el perquè tot i els moltíssims programes i recursos d'investigació científica –universitaris i empresarials– promoguts fins ara, els resultats siguin tan qüestionables.

La resta del document anterior s'ocupa d'explicar "grosso modo" per on es troba la resposta a les preguntes, la "teoria que avui encara no ens podem ni imaginar", i que a la vegada donaria una resposta específica a la "nova tecnologia" que requereixen els Cercadors.

El que també és segur, és que no ens cal invocar a Deu per a trobar la solució tal com fa Sergey BRIN, cofundador de GOOGLE:

"El motor de recerca perfecta seria com la ment de Deu".

La solució està moltíssim més a prop. Està dins nosaltres mateixos, en les nostres mateixes facultats intuïtives. I per a reconèixer-les, calen molts dels coneixements científics d' "occident", una basta interdisciplinarietat, però també aplicada amb les actituds humils i introspectives d' "orient".

Es diu que la solució del futur haurà de ser compatible amb la majoria de sistemes en ús, i més especialment amb els “standards” actuals –o futurs– del mercat, o encara més, amb totes les architectures.

L'ERROR

Ser pragmàtic és una manifestació de supervivència, un instint bàsic i imprescindible, per definició. Si les velocitats de processament es multiplicaven per milions, era obligat aprofitar-ho tal com s'ha fet.

L'error és l'habitual banalització de l'home. Una cosa és ser pragmàtic, l'altra menysprear lo qualitatiu, lo comprensiu o lo intrínsec. Ho explica el principi “zero” de la psicologia, de la pedagogia i de la semiologia, a l'apartat [201] de “Què és la intel·ligència?”: “la psique no va evolucionar per a pensar sinó que per a sentir i per a moure el cos”. Els processos de simplificació conceptual [2187] són una mostra des de l'inici de la formació del pensament. Pensar suposa un esforç, una disciplina, que l'home sempre intenta evitar, llevat que no li quedi més remei.

És el que va passar amb la Intel·ligència artificial. Primer s'intenta –com també es fa amb la “Vida artificial”– emular a Deu*, crear intel·ligència “sintètica”. Després –en lloc de treure conclusions a l'estrepitós fracàs del 1992– es trivialitza reduint la definició d'intel·ligència a estúpids automatismes.

* La religió és una curiosa manifestació humana molt arrelada i no només dins la mateixa religió. La ciència també (com amb el creacionista “origen de l'univers” en lloc de la “materialització de l'univers”), fins i tot la matemàtica també (com amb el creacionista “formalisme” de la lògica, tractat a [3210]) i sens dubta tota la informàtica, on, com s'està veient, les referències a Deu són constants.

L'error va ser el triunfalisme de la velocitat i oblidar-se de seguir preocupant-se per la eficiència, per lo comprensiu i per lo intrínsec. L'error va ser abandonar la recerca de base, tot i que aparentment, pel que es deia, tothom feia recerca. !No!, tothom feia “aplicació”, no “recerca”, dos coses molt diferents. I mentres s'han perdut uns quants anys.

UN AS, UNA CLAU ANGLESÀ, UN PRODUCTE UNIVERSAL

Per tantes exigències d'universalitat –sistemes, “standards”, architectures, plataformes– s'entén l'anterior invocació divina perquè avui no es coneix un altra recurs, per l'error ja esmentat. No es coneix, però hi és.

En llenguatge matemàtic això que es demana té un nom molt clar “Producte universal (dins una categoria d'elements)”. El problema ja l'haurien resolt els matemàtics si només es tractés de números (de fet ja esta resolt des de fa anys amb els “Tensors”), però el problema està en que els “elements” que gestionem aquí no són numèrics, sinó que conceptuals i cognitius.

Aquest “Producte universal” de l'àmbit conceptual-cognitiu, compatible amb tot el que hi ha –o pugui haver–, siguin sistemes, “standards” o architectures, només pot estar dins nosaltres mateixos com ja s'ha dit. I en un lloc anterior a quan el nostra pensament expressa la seva capacitat d'inventar sistemes, “standards” o architectures. I sens dubta, molt abans dels mots i de les pàgines –els papirs, els llibres de GUTENBERG o les anacròniques pàgines Web d'avui– que no representen res del nostra pensament, sinó que són recursos forçats per les limitacions materials dels suports de la representació.

Aquest “Producte universal” és el “Sistema conceptual-cognitiu* (intrínsec i exacta)”, del que a l'inici ja s'ha esmentat la seva estructura (espai, successions, canvis de base, ...), i que s'explica més extensament a “L'estructura del pensament i del coneixement” i als documents referenciats en ell.

* “Grosso modo”, un “Sistema” és un conjunt d'elements interrelacionats. Com un “Coneixement” són dos conceptes relacionats, en conseqüència, un sistema de conceptes genera un conjunt de coneixements, és a dir, un sistema de coneixements. En conseqüència, sistema conceptual i sistema cognitiu és, si fa o no fa, el mateix.

De “Classificacio.pdf” traslado:

¿Quina és la estratègia informacional dels actuals Cercadors d'Internet-WWW?: **cap**. Un pas encara més enrere respecte artificioses estratègies com les referències del “Permuterm Subject Index” de fa més de 50 anys. Un **nul procés de classificació** (què és una estratègia informacional **bàsica**) ¿Com pot ser això possible? Doncs perquè s'indexen gairebé **totes les paraules** de les pàgines Web. Amb paraules l'ordinador no pot entendre i no es pot plantejar **cap estratègia comprensiva** (classificació, ...), només

l'enumerativa d'indexar-ho tot. Aquesta brutal indexació només ha estat possible darrerament, amb un "hardware" d'emmagatzemament de dades de dimensions monstruoses, suportat en una tecnologia de miniaturització impensable fa pocs anys. Però tot i que funcionen, la **servitud** d'aquesta nul·la estratègia comprensiva és clara: qualsevol consulta ens retorna **milers o milions** de pàgines Web candidates. Gairebé tot és "soroll"/ "brossa".

Un exemple, d'entre els milions que es poden trobar, va ser al fer aquesta Web www.sistemaconceptual.org. Per a la seva pàgina secundària "**Proposta**" buscava imatges de galàxies per a il·lustrar un exemple d'una vella demostració sobre el concepte "**Forma**". Veient el poc que trobava amb la paraula "Galàxia" (comuna pel català i el castellà/ espanyol), vaig fer-ho en anglès. El resultat **70 000 000 de pàgines Web!!** (això sí, en només 0.07 segons!!) I el pitjor problema, que primer apareixien infinitats d'empreses (Galaxy technology, Los Angeles galaxy, Multicines Galaxy, ...), cotxes (Ford Galaxy, ...) i qualsevol altre cosa totalment independent d'una "galaxy".

El problema dels Cercadors és:

- **indissociable** del problema de les traduccions inequívokes/ exactes,
- un problema parcial d'un problema genèric, el de la Classificació i recuperació de la informació.

Per això es remet al lector als documents "[Classificació.pdf](#)" i "[Traduccions inequívokes/ exactes](#)".

El problema comença perquè no se sap diferenciar entre "paraula" i "concepte". El concepte "Forma" pot servir d'exemple d'això, amb la **comprensió que s'obté** d'aquest concepte quan es coneix la seva profusa relacionabilitat, aportada pel Sistema conceptual, en contrast amb la nul·la comprensió i gestionabilitat de la "paraula" Forma. Veure el document "[Forma.pdf](#)", del que trasllado:

Amb extrínseques i ambigües "paraules", sense conceptes en una estructura exacta i relacional com aquesta ¿com es pot pretendre afrontar problemes com les **traduccions**, els "**operadors virtuals**", o pitjor, la "**programació lògica**" o la "**intel·ligència artificial**"? És com començar una casa per la teulada (i els resultats estan a la vista). També és d'una irresponsabilitat màxima, pels ingents recursos que s'estan aplicant arreu des de fa anys.

Pel greu problema que els cercadors tenen amb els idiomes, em remeto a una altra de les aplicacions del Sistema conceptual: la desambiguació i les traduccions exactes/ inequívokes. Veure al respecte el document "[Forma.pdf](#)" i "[Traduccions exactes.pdf](#)".

REFLEXIÓ PERSONAL

La prestació d'un servei dona dret a una retribució. Per exemple, un servei podria ser precisament una recerca d'informació, i la retribució podria ser que l'usuari fos destinatari d'uns missatges publicitaris (una retribució indirecta, on el cost directa del servei de la recerca el paga un tercer, diguem-li "Coca-caó", que és qui s'anuncia). Si no s'exerceix el dret a la retribució, implica un regal del servei. Els cercadors d'Internet-WWW ens regalen un servei, el de recerca. Però retribuit o regalat, el servei sempre ha de ser el que hem demanat. Seria inadmissible que si demanem patates ens donin cadires.

Si la intervenció d'un tercer –qui paga el servei– comporta modificar el servei, és una alteració del servei. És el que succeeix quan la informació recercada és modificable en funció dels interessos del tercer. Molts cercadors actuals obtenen part del seu finançament d'aquesta manera: qui els paga, no paga per anunciar-se sinó que per a modificar el resultat de les recerques dels altres.

En un sistema conceptual això no és possible (és a dir, es garanteix que no passi) perquè no existeixen identificadors. Ni empreses, ni marques, ni productes, ..., ni persones, ..., perquè tot això són identificadors. Els identificadors són el nivell 0 (veure [\[.210\]](#)), un nivell previ al pensament*. El sistema conceptual serveix precisament, entre altres moltes aplicacions, per a classificar identificadors amb "**Adreces exactes**". Res a veure, tot un altra món. Veure "[Breves comentaris.pdf](#)" i "[Classificació.pdf](#)".

* En un altra ordre de coses, notem que el Sistema conceptual comença precisament allà on gairebé acaben, on ja no arriben, altres recursos com les "enciclopèdies", que sovint s'han de restringir a biografies, fets històrics, topònims, ... i uns pocs temes elementals. Veure l' "[Aplicació del Sistema conceptual a les Enciclopèdies](#)".