
LARCA: Laboratori d'Algorísmica Relacional, Complexitat i Aprenentatge

Descripció dels objectius, el pla de treball i les activitats per al període d'aplicació del Pla de Recerca i Innovació (2005-2008)

Descriure detalladament quins són els objectius que es pretenen assolir en els propers quatre anys. Aquests, no han de ser únicament objectius científics sinó que també han de fer referència a la planificació del grup amb assoliments específics quantificables i mesurables dins el període de referència (per ex: tesis doctorals llegides, nous projectes obtinguts, patents, contractats postdoctorals, etc.). Cal tenir en compte l'adequació a les orientacions del Pla de Recerca i Innovació 2005-2008 i els criteris d'avaluació enumerats a la resolució de la convocatòria. Els objectius seran valorats *in itinere* al cap de dos anys de la concessió de l'ajut i definitivament al quart i darrer any del període d'aplicació de l'ajut. Especificar el calendari i el pla de treball.

Màxim 4 pàgines en total.

A l'àmbit de recerca propi del nostre grup, sabem per pròpia experiència que no és pas convenient definir objectius temàtics de recerca amb massa precisió per un període de 4 anys. El nostre grup s'enquadra sense dificultat en un àrea temàtica prioritària a nivell català, espanyol i europeu (Tecnologies de l'Informació i les Comunicacions) al qual pretenem continuar. Ara bé, la competitivitat per projectes de recerca a aquesta àrea ja és molt gran, i encara creixerà més; només una vegada ens ha fracassat una sol·licitud de financiació espanyola, però, el projecte integrat i la xarxa d'excel·lència aconseguits al sisé programa marc de la Unió Europea han sigut al cost de més de sis iniciatives similars rebutjades. Això ens mostra que hem d'estar disposats a fer recerca a un seguit d'àmbits i amb un seguit de plans diferenciats, per tal d'aprofitar les oportunitats més adients de convocatòries de projectes, i després desenvolupar aquells pels quals s'ha aconseguit la financiació. Evidentment, a hores d'ara és impossible comprometre'ns a garantir uns objectius de recerca massa específics amb el risc de que la financiació que eventualment es trobi no hi correspongui.

Anàlogament, cada tema de recerca presenta unes possibilitats diferents de publicació i aprofitament dels resultats, i per tant totes les estimacions que oferirem, com s'ens demana, sobre previsió de publicacions depenen també d'aquest factor imprevisible.

Així doncs, els nostres objectius han d'estar guiats per dues consideracions: primer, aquells que corresponen a tasques que, sigui quin sigui el desenvolupament posterior de les sol·licituds de finançament de recerca, constitueixen objectius prou rellevants per al grup com per assegurar-ne la necessitat; segon, aquells que corresponen a tasques ja encetades, potser amb financiació ja obtinguda. Només en tercer lloc, presentem propostes addicionals amb les que volem garantir la capacitat del grup per a adequar-se a canvis a les prioritats de les institucions corresponents, si es dona el cas.

Dins del primer grup d'objectius, hi distingim tres nivells:

a/ nivell intern al grup: coherència i progrés

b/ nivell de relacions amb altres ens de la nostra universitat

c/ nivell de relacions amb el món extern a la nostra universitat

En referència al nivell a/, tenim el contexte següent. Com esmentem a un apartat anterior, una part àmplia d'aquest grup de recerca ha estat enquadrat, a la prèvia estructura de línies de recerca de la UPC, juntament amb una part del grup de nova creació ALBCOM. Amb la recent reorganització, hem apostat per la configuració actual per tres motius: integrar amb nosaltres el grup d'en Glyn Morrill, que pertanyia a altres línies de recerca però que participa amb nosaltres des de fa temps als projectes CICYT; diferenciar temàticament el nostre treball, que té característiques pròpies, del de la resta d'ALBCOM; i, degut al fet de que les prioritats europees més clarament al nostre abast rauen a la línia proactiva de recerca Complex Systems dins del programa FET de ESPRIT (a la qual pertany el projecte integrat DELIS del sisé programa marc al que participem), hem considerat adient augmentar el grup amb la participació d'en Jordi Delgado, expert en aquest àmbit. Tenim, a més, com a clar indicador de progrés dels nostres investigadors, l'objectiu de que els no doctors assoleixin el doctorat.

Podem, doncs, proposar els següents objectius a nivell intern, O1, O2 i O3, amb els respectius indicadors O1.a, O1.b; O2.a, O2.b, O2.c; O3.a, O3.b, O3.c, O3.d:

O1/ Millorar la cooperació, que ja fa temps es dona, entre les persones que treballen principalment a Computational Learning i Data Mining (Gavaldà, Castro, Balcázar) amb les que treballen principalment a Mathematical Linguistics (Morrill, Fadda); volem progressar cap a aquest objectiu: (O1.a) obtenint un nou projecte de CICYT amb participació de Gavaldà i Morrill, entre d'altres participants del grup o de fora d'aquest, i orientat a una tasca comuna a tots dos col·lectius, específicament el modelatge individualitzat de seqüències simbòliques; (O1.b) produint si més no una publicació conjunta al llarg del període sota descripció.

O2/ Involucrar activament als recercaires que procedeixen d'altres àmbits. Es tracta d'en Jordi Delgado, ja esmentat, i d'una persona de la nostra universitat que forma grup de recerca (a la convocatòria SGR) amb un equip d'una altra universitat (en Santiago Boza) i que, si som capaços de desenvolupar una cooperació progressivament major, podria potser passar al nostre en una futura convocatòria, com de fet ja hi pertany a nivell intern de la UPC. Així doncs, en aquest sentit pretenem: (O2.a) produir si més no una publicació conjunta amb en Jordi Delgado al primer període de dos anys, (O2.b) produir una segona al segon període de dos anys, i (O2.c) produir si més no una publicació conjunta amb en Santiago Boza al llarg dels quatre anys. Desitjariem, però no ho podem garantir, que aquestes publicacions es produïssin al contexte d'una activitat finançada.

O3/ Creïxer a un ritme sostenible, en quant a qualificacions personals i en grandària. Específicament, esperem que si més no un dels tres membres del grup que no són doctors a hores d'ara assoleixi el grau de doctor dins dels primers dos anys (O3.a), i que tots tres ho facin dins del període de quatre anys (O3.b). Addicionalment, creiem a la recerca que fem i n'expliquem la seva necessitat, i allò ens motiva per fer esforços en fer créixer el grup, per tal de desenvolupar més i millor les tasques que identifiquem com a objectius científics. Creiem que un creixement sostenible es pot definir com la incorporació de nous doctorands, en quantitat entre 2 i 5 (O3.c), amb suport financer assignat al grup sempre que es pugui. Si convingués per al pla de treball de cada doctorand, aquest podria desenvolupar la seva tasca repartint el seu temps de doctorat entre el nostre grup

i altres grups de recerca, i si escau assolir la menció de Doctor Europeu. La nostra participació a la sol.licitud Marie Curie APPLY, sota avaluació, es dirigeix en aquest sentit, així com una nova proposta actualment en preparació mitjançant la qual esperem finançar l'estudiant de doctorat Albert Bifet, qui s'incorporaria al grup. Esperem continuar també amb la possibilitat, que ja hem emprat amb èxit en diverses oportunitats, de comptar amb ajuts puntuals d'estudiants de primer i segon cicle que, per exemple fent els seus projectes de fi de carrera o treballs dirigits amb nosaltres, o bé treballant finançats per projectes europeus, o bé mitjançant beques de col.laboració, ens fan cops de mà a la nostra tasca i aprenen nous coneixements útils, moltes vegades connectats amb el món industrial.

Cal dir també que, amb aquest creixement, preveïem dues necessitats addicionals cap als anys 3 i 4: primer, la de dotar el grup d'una estructura més elaborada, que amb la grandària actual no veiem necessària però que caldrà per créixer adequadament, i que el.laborarem a principis del tercer any (O3.d); i, segon, la de disposar de personal de suport a la recerca a temps parcial, que a hores d'ara no ens és encara imprescindible.

Al respecte de les nostres relacions amb altres ens de la nostra universitat, cal esmentar que ja es dona el cas de què membres d'altres grups de recerca de la UPC (i en el seu cas, d'altres universitats) participin als mateixos projectes de recerca, i, de fet, ja existeixen publicacions, o bé són de propera aparició, conjuntes amb membres d'ALBCOM i TALP; així doncs, com a objectiu a aquest apartat,

O4/ esperem continuar una fructífera cooperació, en particular, amb ALBCOM i TALP, que esperem es concreti en, si més no, (indicadors O4.a i O4.b respectivament) una nova publicació conjunta amb si més no una persona de cadascú d'aquests grups. El contexte idoni és el projecte Trangram de CICYT, on participen membres del TALP, i que és un subprojecte d'un projecte coordinat amb altres dos subprojectes, un dels quals (GRAMMARS) és liderat per ALBCOM, i reflecteix un enfoc multidisciplinar a models basats en grafs per estructures formades per quantitats inmensos d'elements individuals. Estudiarem la possibilitat d'estructurar amb ells entitats de recerca més voluminoses, si convé per motius científics.

O5/ Existeix, però, una altra cooperació que tot just s'enceta a hores d'ara, i de la qual el contexte legal no és encara prou estable per presentar objectius específics: el canvi de plans d'estudi per tal d'adaptar les universitats espanyoles al marc de Bolònia, i que indirectament ens afecta degut a la substitució conjunta dels actuals segons

cicles i cursos de doctorat per estudis de postgrau, així com l'evolució de cicles curts i llargs de l'actualitat a estudis de grau, afectant així fortament els estudis de doctorat. El nostre grup s'ha ofert a cooperar amb l'EPSEVG, un centre que fins fa poc només oferia estudis de cicle curt, no només per la creació dels estudis de grau d'Informàtica d'acord amb la nova legislació, sinó també d'estudis de postgrau que puguin conduir a la recerca i a un títol de doctor. A hores d'escriure aquest document, tantmateix, no hi ha decisions finals preses pel Ministeri al

respecte d'aquesta evolució, i només ens podem comprometre a (indicador O5.a) oferir i, si escau, liderar una proposta de títol de postgrau orientat a la recerca a la EPSEVG, que volem que sigui el més compatible possible amb els estudis que ofereixin altres nodes de la nostra xarxa d'excel·lència PASCAL del sisé programa marc. Creiem probable que aquesta cooperació amb l'EPSEVG es formalitzi mitjançant la creació d'un subgrup del nostre grup de recerca que quedi vinculat expressament a l'EPSEVG, decisió que, en el seu cas, es recolliria amb l'estructuració del grup prevista per a començament de l'any 3, com ja queda dit a dalt.

O6/ Per finalitzar el primer àmbit, les relacions amb el món extern a la nostra universitat actualment es concreten en un ampli ventall de contactes universitaris, majoritàriament però no solament europeus. Volem mantenir i reforçar aquestes contactes amb una política mesurada de visitants, que avaluem en tres mesos de visitant l'any. Tantmateix, les connexions locals, més enllà de la UPC, no són prou fortes. Desitjem evolucionar cap a una millor configuració. Ens proposem, primer, establir millors contactes amb grups de recerca d'altres universitats de Catalunya, que es traduïrien en (indicador O6.a) si més no, una nova iniciativa de recerca a engegar durant el període de quatre anys i amb participació de si més no un grup més, d'una universitat catalana diferent de la nostra, amb continguts que encara no podem fixar; segon, (indicador O6.b) incrementar la nostra participació a les activitats de la xarxa espanyola MINDATOS, bé aportant-hi publicacions, bé organitzant una jornada de treball conjunt, bé assumint el lideratge d'iniciatives concretes al seu si. En tercer lloc, ens proposem treballar per remediare una de les mancances més obvies del nostre grup, és a dir, els contactes directes amb el món empresarial: durant els primers dos anys establirem contactes amb empreses per tal de, no més tard del tercer any, (indicador O6.c) preparar i presentar una iniciativa de sol·licitud de projecte PROFIT. Com a oportunitat de millorar la nostra visibilitat a diferents nivells, (O6.d) organitzarem al 2006 (simultàneament) els congressos Algorithmic Learning Theory i Discovery Science, i (O6.e) el "workshop" d'àmbit estatal Learning'06, així com (O6.f) International Conference in Database Theory 2007 conjuntament amb membres d'ALBCOM. Finalment, sense gosar concretar un indicador específic, hi participarem a tota iniciativa de finançament europeu de recerca que ens sigui proposada i no distorsioni els nostres interessos científics, com sempre hem fet.

En referència a tasques per les quals ja disposem de financiació, generalment parcial i que en alguns casos esperem complementar o prorrogar amb nova financiació ja demanada i actualment sota avaluació, volem esmentar temes concrets de treball als que pretenem contribuir amb publicacions de nivell internacional.

1/ Dins del camp de l'aprenentatge computacional, i amb finançament parcial de l'Acció Integrada amb Alemanya, esperem contribuir amb una anàlisi addicional del cas d'aprenentatge amb algorismes del model de la computació quàntica

2/ Dins del mateix camp, millores i implementacions dels algorismes d'optimització per a mètodes de vectors suport (projecte Trangram de CICYT,

ja atorgat i a punt de començar, i projecte FUNCOPT demanat a l'ESF i encara sota avaluació)

3/ Dins del mateix camp, avenços als algorismes actius basats en interseccions booleanes (projecte Trangram ja esmentat)

4/ Dins del camp de Data Mining, avenços conceptuals sobre estructures tancades en seqüències i d'altres dades estructurades (projecte Moisès de CICYT en curs, projecte integrat europeu DELIS)

5/ Dins del mateix camp, adaptació, millora, implementació, i, si escau, transferència al sector industrial dels algorismes de mostreig adaptatiu per modelitzar seqüències no estables en el temps (projecte integrat europeu DELIS, nova financiació ja demanada, i nova financiació que demanarem properament)

6/ A un camp conceptualment més elevat i, per tant, més difícil d'etiquetar, estem buscant un enfoc genèric al procés de modelatge de les restriccions presents a dades relacionals extensionals, que pretenem basar sobre nocions d'Anàlisi de Conceptes Formals (finançament de la xarxa PASCAL)

7/ També a un nivell molt abstracte, una generalització de la noció de 2-estructures sobre grafs que es pugui aplicar a l'anàlisi de dades relacionals mitjançant la descomposició de grafs de Gaifman (projecte Moisès de CICYT)

8/ Creiem que podem assolir un ventall d'aplicacions noves de la generalització esmentada en el punt anterior, en particular en referència a nocions de gramàtiques dites "mildly context-sensitive", i potser també aplicacions a modelatge de "clustering" i dels algorismes existents d'aprenentatge actiu mitjançant preguntes d'equivalència i pertinença

9/ Avenços a la modelització de fenòmens de discontinuïtat presents a moltes llengües naturals, d'una manera anàloga al modelatge de seqüències (projecte Moisès de CICYT)

Aquests temes concrets ens proporcionaran les tres tesis doctorals esmentades més amunt (en particular, els temes 4, 6 i 9); i un seguit de publicacions, que procurarem siguin del màxim nivell internacional, i amb les que creiem que podem assolir els següents indicadors per als primers dos anys: (O7.a) si més no, dues publicacions en revistes internacionals contenint resultats majoritàriament ja obtinguts a hores d'ara però només anunciats en forma preliminar a congressos i workshops; (O7.b) si més no, cinc publicacions a congressos internacionals descrivint avenços completament nous; i, per als dos anys següents, (O7.c) tres publicacions addicionals a revistes internacionals i (O7.d) quatre publicacions a congressos internacionals.

Aquests temes, amb probabilitat elevada, no seran adients per als anys 3 i 4, o només ho seran en part. Els enfoc que ara són més "visionaris" potser rebin una continuació molt més concreta. Esperem que, amb si més no un dels temes (però, qui sap quin!, en tenen la major probabilitat 2, 4 i 5), tindrem la

sort de trobar coneixement que admeti transferència al teixit industrial, i buscarem un marc adient per desenvolupar-lo, potser via projectes PROFIT. I de segur que molt aviat tindrem, a sobre de la taula, noves preguntes que resoldre; a tall d'exemple, hem demanat finançament Marie Curie amb la xarxa APPLY per tal d'estudiar problemes de "metadata imputation", amb potencialitat d'aplicació pràctica a la web semàntica, mitjançant tècniques de nuclis que complementarien els resultats de la línia 2 esmentada. També hem començat contactes amb el Laboratori d'Aplicacions Bioacústiques de la EPSEVG, on esmerçarem esforços per tal d'identificar, si és possible, vies de cooperació. Ens agradaria que, en part, algunes d'aquestes noves preguntes que ens plantejarem tornessin a caure a l'àmbit de la Complexitat Computacional, una línia de recerca que ha perdut prioritat a Europa (com ho mostren els successius fracassos de les propostes de projectes europeus als què hem participat) i que ens esforçarem en recuperar com a prioritat de recerca, ara amb el nou, recent argument de que el Clay Mathematical Institute de Boston (Estats Units) ha inclòs el problema central de la Complexitat Computacional (P versus NP, amb variants del qual hem treballat fort fa anys quan encara es rebia finançament per aquesta tasca a Europa) com a un dels set (7) problemes matemàtics del mil.leni i, com a tal, dotat d'un premi d'un milió de dòlars per a qui ho resolgui (veure www.claymath.org/millennium). A tal fi, participarem a noves iniciatives que esperem sorgeixin de la propera reunió Computability in Europe a Amsterdam al mes de juny.

De la mateixa manera, podríem participar, aportant-hi les nostres capacitats, a iniciatives on el modelatge d'informació implícita, i el seu estudi des del punt de vista de la Teoria de la Complexitat, sigui de relleu; a tall d'exemple, hi ha fortes connexions entre el nostre treball i les tasques desenvolupades al camp d'Information Retrieval; alguns de les nostres anàlisis en aprenentatge actiu podrien trobar aplicació al camp de Recommender Systems, fortament desenvolupat als Estats Units (i amb una immensa inversió en recerca procedent del sector privat) i molt subdesenvolupat a Europa; o, finalment, a àmbits on sigui precís trobar representacions comprensibles d'estructures compostes de milions d'elements, que, com ja s'ha dit, és l'objectiu guia del projecte GRAMMARS, del qual Trigram és un subprojecte, i que poden venir motivades per situacions on sigui precisa l'aplicació de moltíssimes quantitats d'elements petitíssims en cooperació, com ara, possiblement, a iniciatives del camp de la NanoEnginyeria.