

“Generació de metadata semàntica mitjançant jocs online”

Enginyeria Tècnica Informàtica de Sistemes
Treball Final de Carrera

Sergi Soler i Segura
Desembre de 2008

1 Índex

1	Índex	2
2	Presentació	4
3	Introducció.....	4
3.1	Abast del projecte	5
4	La web semàntica	6
4.1	La web semàntica. Breu ressenya històrica	6
4.2	La Web dels Documents i la Web de les Dades	7
4.3	Metadades	8
4.4	RDF. Resource Description Framework	8
4.4.1	RDFS. RDF-Schema	9
4.4.2	RDFa. Resource Description Framework in Atributes	10
4.5	OWL. Web Ontology Language	11
4.6	Problemàtica en la implantació de la web semàntica	11
4.6.1	Problemàtica associada a la anotació de recursos web.....	12
5	Generació de coneixement mitjançant jocs online	13
5.1	Breu ressenya històrica	13
5.2	Human Computation.....	13
5.3	Games with a purpose (GWAP).....	14
5.3.1	ESP Game	14
5.3.2	Verbosity	15
5.3.3	Característiques de disseny	15
5.3.4	Tipologies de joc	16
5.3.5	Resultats	17
5.4	Anotació de recursos web mitjançant jocs online	17
5.5	Exemples de jocs per la generació de coneixement.....	18
5.5.1	Ontogame	18
5.5.2	OntoPair: Towards a Collaborative Game for Building OWL-Based Ontologies.....	19
5.5.3	Listen Game - A game-based approach for collecting semantic annotations of music	19
5.5.4	FACTory.....	20
5.5.5	Altres exemples.....	21
5.6	Valoració de la situació actual	21
6	Joc online per l'anotació de blogs	22
6.1	Weblogs.....	22
6.1.1	Subscripció a blogs	22
6.2	Anotació de blogs	23
6.2.1	Anotació automàtica.....	24
6.3	Criteris de disseny	24
6.4	Ontologies útils per l'anotació de blogs	26
7	Prototip – Requisits	27
7.1	Finalitat del joc.....	27
7.2	Criteris de disseny	27
7.3	Funcionament del joc.....	28
7.4	Nota d'aclariment.....	28
8	Prototip – Anàlisi	29

8.1	Modes de Joc	29
8.1.1	Mode de joc Two-Players.....	29
8.1.2	Mode de joc Single-Player	29
8.2	Actors	29
8.2.1	Visitant	29
8.2.2	Usuari.....	29
8.2.3	Jugador	29
8.2.4	Sistema	30
8.2.5	Administrador	30
8.2.6	Diagrama d'estats pels visitants del web.....	30
8.3	Casos d'ús.....	31
8.3.1	Diagrama de casos d'ús.....	31
8.3.2	Registre d'Usuari.....	31
8.3.3	LoginUsuari	32
8.3.4	Gestió del perfil d'Usuari	32
8.3.5	Inici Partida	32
8.3.6	Fer una jugada	32
8.3.7	Processar una jugada	33
8.3.8	Servir un 'post'.....	33
8.3.9	Activar mode Single-player	34
8.3.10	Finalitzar la partida.....	34
8.3.11	Executar Anotació	34
8.4	Algorisme d'inici de partida.....	35
8.5	Algorisme de detecció d'abandonament d'un jugador	36
9	Prototip – Disseny	37
9.1	Variacions respecte l'etapa d'anàlisi.....	37
9.2	Funcionament del joc.....	37
9.3	Diagrama de classes	38
9.4	Inici de partida	39
9.5	Servir Round.....	40
9.6	Funcionalitats del sistema	41
9.6.1	Servir post	41
9.6.2	Incorporació de posts a la DB	41
9.6.3	Sistema de puntuació.....	41
9.7	Persistència	42
9.7.1	Definició d'entitats (taules)	42
10	Prototip – Implementació	44
10.1	Arquitectura	44
10.2	Codificació	45
10.3	Components	45
10.3.1	SimplePie	45
10.3.2	jQuery.....	45
10.3.3	RDFlib	45
10.3.4	MySQL	46
10.4	Fitxer RDF	46
11	Conclusions	47

2 Presentació

El present document presenta la memòria del **Treball Final de Carrera (TFC)** dels estudis d'Enginyeria Tècnica Informàtica de Sistemes de la Universitat Oberta de Catalunya – UOC.

El TFC es desenvolupa en l'àmbit general de la **Web Semàntica**, i concretament es centra en l'estudi dels mètodes d'anotació automàtica de recursos web mitjançant jocs online. En base a aquest estudi, s'ha dissenyat i implementat un joc online que permet l'anotació automàtica de weblogs.

3 Introducció

Aquest TFC estudia com amb la utilització de jocs online es podria resoldre, o si més no contribuir a resoldre, un dels problemes que fan que la web semàntica avui dia encara no sigui una realitat, per bé que va ser postulada l'any 1999 per Tim Berners-Lee a la seva obra "Weaving the Web". El problema a resoldre és l'anotació de continguts web i la solució que es proposa és aplicar el concepte de 'Human Computation' a un joc online, de manera que els jugadors a base de partides, solucionin el problema plantejat.

La web semàntica (popularment també se l'ha anomenada Web 3.0) és una extensió de la web tradicional que coneixem avui dia, que permet que les màquines puguin entendre la semàntica dels continguts web que es troben a Internet.

Per tal que la web semàntica sigui una realitat hi ha diversos aspectes tècnics que cal abordar: construcció d'ontologies, alineació d'ontologies i anotació de recursos, essent aquest últim, l'objecte d'aquest TFC.

La web semàntica requereix que els continguts web incorporin una sèrie d'anotacions que permetin als ordinadors entendre'n el significat. El procés de crear aquestes anotacions és el que anomenem l'anotació de continguts, i aquest es podria fer en el moment en que es desenvolupen els continguts. El problema però, és que actualment, la majoria de professionals del sector no tenen els coneixements necessaris per fer-ho, i a més, hi ha gran quantitat de continguts que els generen usuaris que no són professionals del sector, com són per exemple, tota la comunitat de bloggers. Per altre banda, ens topem també amb el fet que la gran majoria de continguts web actuals no disposen d'aquestes anotacions. Així doncs, la falta d'aquestes anotacions s'ha convertit en un coll d'ampolla a l'hora de fer realitat la web semàntica.

La feina de crear aquestes anotacions, avui dia, és una tasca que requereix la intervenció humana, i el cost econòmic que això suposa fa que sigui inviable si tenim en compte la gran quantitat de recursos que hi ha per anotar, i la naturalesa canviant d'aquests. És per això que es fa necessari trobar mètodes d'anotació automàtics.

Actualment, sembla que la anotació automàtica mitjançant jocs online, seguint el paradigma "Games with a purpose" de Luis von Ahn, pot ser una bona solució al

problema exposat. Per una banda els participants en el joc ho fan de manera desinteressada, la qual cosa soluciona el problema del cost econòmic associat a l'anotació manual, i per altre banda, els usuaris potencials del joc són tota la comunitat d'internautes, amb la qual cosa el procés d'anotació de tota la web, és factible. A més, no cal que els usuaris tinguin coneixements de la web semàntica, quedant el procés d'anotació amagat darrere del joc.

3.1 Abast del projecte

El present document està estructurat seguint les etapes de desenvolupament per les que ha passat el projecte, que a la vegada marquen quins són els objectius del projecte.

A l'apartat 'La web semàntica' es veuen els fonaments de la web semàntica i s'estudia quina és la problemàtica associada a l'anotació de continguts web. A l'apartat 'Generació de coneixement mitjançant jocs online', s'estudia el concepte de 'Human Computation' i la seva aplicació als jocs online establint així el que es coneix com el paradigma 'Games with a purpose' i es veuen les diferents iniciatives que hi ha actualment a la web per l'anotació de continguts. A l'apartat 'Joc online per l'anotació de blogs' es veuen les característiques bàsiques a tenir presents pel desenvolupament d'un joc d'aquestes característiques. Finalment, als últims capítols es veuen les fase d'anàlisi, disseny i implementació d'un joc online, que té per objectiu la anotació dels posts de blogs on la temàtica sigui la poesia.

En el desenvolupament d'aquest TFC, intervenen diferents àrees de l'Enginyeria Informàtica, a la vegada que calen conceptes del camp de la gestió i representació del coneixement, com són, des de la pròpia web semàntica a conceptes com 'Human Computation', 'Collective Intelligence' o 'Ontologia'. Per l'anàlisi, disseny i implementació del joc calen coneixement de tecnologies web (HTML, Javascript, XML, RSS), així com de Enginyeria de Programari, Base de Dades (MySQL), i Programació (PHP, Python).

4 La web semàntica

Aquest apartat té per objecte entendre què és la web semàntica, quins són els principis sobre els que es fonamenta i com es materialitza. En cap cas pretén ser una explicació formal i exhaustiva dels conceptes que es presenten, ni cap manual dels estàndard que es descriuen, ja que això queda fora de l'abast del projecte.

4.1 La web semàntica. Breu ressenya històrica

Donem una sèrie de dates per tal de centrar històricament el discurs. El World Wide Web (informalment, la web) té els seus orígens al voltant de l'any 1990, essent-ne el creador Tim Berners-Lee. L'any 1999, el mateix Berners-Lee va postular la web semàntica en el seu llibre titulat "Weaving the Web", al qual correspon la següent cita:

"I have a dream for the Web [in which computers] become capable of analyzing all the data on the Web – the content, links, and transactions between people and computers. A 'Semantic Web', which should make this possible, has yet to emerge, but when it does, the day-to-day mechanisms of trade, bureaucracy and our daily lives will be handled by machines talking to machines. The 'intelligent agents' people have touted for ages will finally materialize."

Tim Berners-Lee, Weaving the Web, 1999

Per contextualitzar aquestes dates, podem afegir que Google es va fundar l'any 1998¹, i que segons alguns autors, marca l'inici del que finalment s'ha acabat anomenant la Web 2.0, on Wikipedia (fundació l'any 2001) i YouTube (fundació l'any 2005) en són dos dels màxims exponents. En aquest sentit, podem dir que la web semàntica és tant antiga com Google, o com la Web 2.0, si més no a nivell conceptual. Per acabar, ressenyar que el gener de 2007 es va llençar la primera versió de DBpedia.org, una de les aplicacions importants que hi ha actualment basada en l'ús de les tecnologies de la web semàntica.

¹ <http://www.google.com/corporate/history.html>

4.2 La Web dels Documents i la Web de les Dades

El terme Web de les Dades està essent utilitzat per diferents autors, com el mateix Berners-Lee¹, com a sinònim de la web semàntica, en contraposició a la web que coneixem avui dia, a la qual ens hi referim com a la Web dels Documents. Aquesta nomenclatura però, no només és per distingir les dues webs, sinó que també fa referència a les característiques fonamentals que les diferencia.

Per copsar les diferències, agafem com a exemple els continguts d'una entrada d'un blog que podem trobar a la URL <http://seleccipoetica.blogspot.com/2008/12/mentre-ella-dorm.html>

POEMES DESAPAMATS

Primer premi "Hilari d'Arenys" 1978

Joaquim Soler i Ferret

MENTRE ELLA DORM

Mentre ella dorm
li faig poemes
perquè la nit no se me la mengi

Mentre ella dorm
li parlo baixet
perquè també li vull entrar pel somni

Mentre ella dorm
li pinto el vestit
perquè demà estigui més bella.

Mentre ella dorm
m'apropro al no res
de ser sense ella.

Molt probablement, qualssevol lector normal i corrent, en front d'aquest text serà capaç de dir a quin gènere literari correspon, qui és l'autor i quin és el títol, entre altres dades que es poden extreure. Ara bé, totes aquestes deduccions que una persona sap fer, un ordinador en el context de la web dels documents, no les sap fer. És a dir, si a un cercador web com pugui ser Google fem la cerca "Qui és l'autor del poema titulat Mentre ella dorm", no obtenim una resposta satisfactòria.

A la web dels documents, els ordinadors només perceben els continguts web com un conjunt de text, el qual per ells no té cap significat. Donat aquest munt de text, no saben distingir quin és el conjunt de lletres que fa referència a l'autor i quin fa referència al poema.

¹ <http://www.technologyreview.com/infotech/18451/?a=f>

A la web de les dades (la web semàntica) si que és possible que un ordinador sàpiga respondre a la qüestió de qui és l'autor del poema 'Mentre ella dorm' plantejada anteriorment. De fet, si els continguts estan degudament anotats, un cercador semàntic, a la consulta "Poemes d'amor" hauria de retornar el poema del nostre exemple, per bé que la paraula 'amor' no surt en cap part del text.

La web dels documents, on el format estàndard és l'HTML, està pensada per ser interpretada per les persones, mentre que la web semàntica està pensada per ser interpretada per les màquines i utilitzem l'estàndard RDF per construir-ne els seus documents. Aquests documents són els que expressen la semàntica associada al recurs web al qual es refereixin, i són els que han de permetre als cercadors com els de l'exemple anterior, respondre les qüestions que plantejàvem. En aquest sentit, cal destacar que la web dels documents i la web de les dades no són excloents, sinó que conviuen les dues a la vegada complint cadascuna la seva funció.

4.3 Metadades

Per entendre què és i com funciona la web semàntica, cal entendre que són les metadades, ja que aquestes són la base del seu funcionament.

Les metadades són dades sobre les dades. A l'apartat anterior s'ha vist que els documents contenen informació, i s'ha vist també com podem treure informació sobre aquesta informació. Reprenent l'exemple anterior, hem vist que hi ha una informació associada al document (gènere literari, autor, títol), la qual no forma part del contingut del document en si. Aquesta informació són les metadades.

Seguidament veurem com amb el llenguatge RDF fent us d'aquestes metadades, aconseguim fer el salt de la web dels documents a la web de les dades.

4.4 RDF. Resource Description Framework

RDF és una Recomanació W3C des de l'any 1999 i és una de les tecnologies centrals sobre la que es fonamenta la web semàntica.

RDF és un llenguatge que permet representar la informació dels recursos que trobem al World Wide Web, i que es basa en fer sentències (que en RDF s'anomenen *triples*) sobre aquests recursos, seguint l'estructura subjecte – predicat - objecte. El subjecte representa el recurs que volem descriure, el predicat explica la propietat que volem descriure del recurs, i l'objecte és el valor que pren el subjecte per la propietat especificada.

<u>Joaquim Soler i Ferret</u>	<u>és autor de</u>	<u>'Mentre ella dorm'</u>
subjecte	predicat	objecte

El subjecte i el predicat, donat que són recursos web¹, venen representats per URIs, i l'objecte pot ser un recurs, o bé un literal.

¹ [http://en.wikipedia.org/wiki/Resource_\(Web\)](http://en.wikipedia.org/wiki/Resource_(Web))

RDF és un model de representació abstracte, el que implica que no hi ha una sintaxi predefinida per construir-ne les sentències. De fet, actualment hi ha diversos formats per serialitzar les sentències RDF, com són: RDF/XML, Notation3, Turtle o N-Triples. Actualment un dels formats més importants és RDF/XML, el qual utilitza XML com a format de serialització. La sentència anterior expressada en RDF/XML quedaria de la següent forma:

```
<rdf:RDF
  xmlns:rdf="http://www.w3.org/1999/02/22-rdf-syntax-ns#"
  xmlns:ns1="http://exemple.org/conceptes#">

  <rdf:Description rdf:about="http://www.escriptors.cat/autors/solerjo/index.html">
    <ns1:es_autor>Mentre ella dorm</ns1:es_autor>
  </rdf:Description>
</rdf:RDF>
```

A l'exemple podem veure com el subjecte (Joaquim Soler i Ferret) ve identificat per un recurs web amb la URI <http://www.escriptors.cat/autors/solerjo/index.html> el qual identifica el subjecte de manera unívoca. Observem que el predicat, per la mateixa raó, també està identificat per una URI, en aquest cas: http://exemple.org/conceptes#es_autor

A partir de la construcció XML anterior i de l'ús de vocabularis comuns globalment acceptats, podem veure com s'obre la porta al fet que un ordinador pugui treballar amb metadades i per tant sàpiga quin és el significat de les dades que acompanyen a aquestes metadades. És a dir, un ordinador sap processar fitxers XML i per tant podrà llegir les sentències. A més les sabrà interpretar ja que aquestes sentències venen expressades amb URIs úniques.

Un conjunt de sentències RDF on unes enllacen amb les altres, formen un graf dirigit etiquetat, de manera semblant a com a la web dels documents, els enllaços d'hipertext entre documents també formen un graf. En aquest sentit, ha sorgit el terme Giant Global Graph¹ (GGG) per referir-se a la web de les dades, en contraposició a les WWW que defineixen la web dels documents.

4.4.1 RDFS. RDF-Schema

RDFS és una Recomanació W3C des de l'any 2004, i és un llenguatge per la representació del coneixement, el qual proporciona els elements bàsics per descriure ontologies.

Una **ontologia** és una representació formal d'un conjunt de conceptes referents a un domini de coneixement específic. Les ontologies en el context RDF també s'anomenen vocabularis.

Així doncs, RDFS ens permet descriure els conceptes que intervenen en un domini de coneixement i establir-ne les propietats, a la vegada que permet crear jerarquies entre aquestes i relacions entre ells. Amb RDFS creem els vocabularis que ens serviran per crear sentències RDF.

¹ <http://dig.csail.mit.edu/breadcrumbs/node/215>

Com a exemple, podríem pensar en una ontologia sobre el món del futbol. En aquesta, definiríem una classe 'Jugador', i definiríem una classe 'Porter' la qual seria una subclasse de 'Jugador', ja que tot porter és també jugador. La definició i posterior acceptació d'aquest vocabulari per part de tota una comunitat, comporta que d'una sentència del tipus "El senyor X és un gran Porter" un membre de la comunitat pugui raonar que 'X és un Jugador'.

Resumint, RDF ens permet crear sentències sobre recursos web, de les quals se'n pot interpretar tota la semàntica si es fa ús de vocabularis globalment acceptats per la comunitat d'usuaris.

4.4.2 RDFa. Resource Description Framework in Attributes

Un aspecte que encara no hem vist, és on s'emmagatzemen les sentències RDF. Hi ha diverses possibilitats. Una de les més senzilles consisteix en crear un fitxer de text amb extensió rdf, el qual només conté les sentències que volem guardar. Altres solucions més complexes, són la persistència en bases de dades relacionals, o en bases de dades natives RDF, anomenades TripleStore.

En qualsevol cas, a les solucions plantejades (fitxer rdf, base de dades), les metadades van a part de les dades. És a dir, estem creant sentències sobre algun recurs web, i aquestes les estem guardant separatament del recurs al qual fem referència.

RDFa (Recomanació W3C des de l'octubre de 2008) és útil per quan volem guardar les dades conjuntament amb les sentències RDF sobre aquestes dades, i es fonamenta en encastar les sentències dins dels documents XHTML, en forma d'atributs dels tags XHTML

En el següent exemple, podem veure un document XHTML abans i després d'haver-li incorporat RDFa

XHTNL

```
<div>
  <div>
    <h2>Mentre ella dorm</h2>
    <h3>Joaquim Soler i Ferret</h3>
  </div>
</div>
```

XHTNL + RDFa

```
<div xmlns:dc="http://purl.org/dc/elements/1.1/">
  <div about="http://seleccio poetica.blogspot.com/2008/12/mentre-ella-dorm.html">
    <h2 property="dc:title"> Mentre ella dorm </h2>
    <h3 property="dc:creator"> Joaquim Soler i Ferret </h3>
  </div>
</div>
```

En aquest exemple diem que estem fent servir el namespace <http://purl.org/dc/elements/1.1/> que identifiquem amb el prefix `dc`. Amb l'atribut `about` identifiquem el recurs que fa de subjecte de la sentència, i amb l'atribut `property` identifiquem el predicat.

A l'exemple tenim dues sentències RDF:

1. <http://selecciopoetica.blogspot.com/2008/12/mentre-ella-dorm.html> té per títol "Mentre ella dorm"
2. <http://selecciopoetica.blogspot.com/2008/12/mentre-ella-dorm.html> té per autor "Joaquim Soler i Ferret"

4.5 OWL. Web Ontology Language

Si bé aquest TFC no fa ús de OWL, si que en aquest apartat se'n fa esment donat que també és una de les tecnologies centrals en la web semàntica. OWL és una Recomanació W3C des de l'any 2004.

A l'apartat sobre RDF, hem definit el que era un ontologia i hem vist que RDFS, és un llenguatge per la definició d'ontologies. OWL també és un llenguatge per la definició d'ontologies, però aquest té un nivell d'expressivitat molt superior que el de RDFS.

OWL està compost per tres subllenguatges: OWL Lite, OWL DL, i OWL Full els quals es diferencien pel grau d'expressivitat

4.6 Problemàtica en la implantació de la web semàntica

Tal com s'ha pogut veure en els apartats anteriors, la web semàntica, tant a nivell conceptual com tecnològic, és ja una realitat. Hem vist que les tecnologies clau que hi intervenen estan ja elevades a la categoria d'estàndard (Recomanacions W3C), la qual cosa vol dir que estan suficientment provades i que tenen el nivell de maduresa necessari per afrontar el repte tecnològic.

Per altre banda però, com a usuaris de la web, constatem que la web semàntica encara no és una realitat. Només cal parar atenció a una de les aplicacions web més importants que existeixen avui dia, com és Google. Aquest cercador, treballa a nivell textual; treballa amb el que hem anomenat la web dels documents, i la web de les dades a hores d'ara no la sap interpretar. El problema però, no és que no es sàpiguen interpretar les dades, sinó que les dades encara no estan en el format adequat per poder ser interpretades a nivell semàntic. Podríem concloure doncs, que la web semàntica existeix, però encara no està implantada en la quotidianitat dels usuaris d'Internet.

Per tal que la web semàntica sigui una realitat a nivell d'usuari, hi ha diversos aspectes tècnics que cal abordar. Aquests són: construcció d'ontologies, alineació d'ontologies i anotació de recursos.

Construcció d'ontologies: Cal construir els vocabularis que ens permetin fer la descripció dels recursos, i cal que aquests vocabularis tinguin un nivell de maduresa tal, que faci que la comunitat d'usuaris els accepti i en faci un ús regular, arribant a convertir-se en estàndards de facto.

Alineació d'ontologies: Freqüentment es pot donar el cas que un mateix terme o concepte es defineixi en ontologies diferents. Això provocarà ontologies que es

solapen degut a termes que es corresponen. Cal crear correspondències per tal de donar consistència al mapa ontològic de la xarxa

Anotació de recursos: Hem vist que la web semàntica es fonamenta en descriure els recursos web mitjançant sentències RDF. Aquest procés de construcció de les sentències RDF és el que coneixem com a anotació de recursos o anotació de continguts. El problema que es planteja a l'hora de l'anotació de recursos, és determinar el qui, el quan i el com es fa aquesta tasca.

Respecte als tres problemes plantejats, aquest TFC es centra en donar una possible solució al problema de la anotació de recursos. Els altres dos queden fora de l'abast del projecte.

4.6.1 Problemàtica associada a la anotació de recursos web

La tasca d'anotació de recursos, és una feina que actualment no s'ha pogut automatitzar completament, i requereix de la intervenció humana. A més, per anotar correctament cal que la persona tingui coneixements sobre la matèria.

Arribats en aquest punt, el que cal veure és sota quines circumstàncies les persones estan disposades a fer aquesta la tasca, la qual es planteja com una feina repetitiva i pesada de fer. Podem enumerar una sèrie de motivacions que podrien portar a una persona a fer aquesta tasca:

1. Per altruisme: Un bon exemple d'usuaris altruistes que han dedicat les seves hores a generar continguts sense cap remuneració a canvi, el tenim amb la Wikipedia. El que passa però, és que sovint al darrere de l'altruisme hi ha motivacions més fortes, com el reconeixement social, reputació professional o fama, i per una tasca tant feixuga com la que ens plantejem, difícilment cap usuari pot trobar-hi cap dels motivacions que hem citat.
2. Per un incentiu econòmic: Aquesta tampoc resulta una bona solució. Si tenim en compte la gran quantitat de recursos web existents, veiem que un iniciativa privada destinada a anotar tots aquests recursos tindria un cost que la fa inviable, a més, cal tenir present que la feina d'anotació no només s'ha de fer la primera vegada que es processa el recurs en qüestió, ja que la naturalesa canviant d'aquests continguts, obliga a haver d'estar-los anotant cada vegada que es realitza algun canvi.
3. Pagar un peatge per ús: Cada vegada que un usuari vol fer servir algun recurs semàntic, cal que primer aportí alguna contribució a la comunitat. Aquesta opció si bé no hauríem de descartar-la, cal tenir en compte que es podria donar el cas que els usuaris arribessin fins i tot a rebutjar l'accés a aquests recursos degut a que cada vegada se'ls interrompria la seva experiència d'usuari per fer-los pagar el peatge.

La solució que es proposa en aquest TFC és amagar el procés d'anotació al darrere d'un joc online, de manera que el que està fent el jugador sense adonar-se'n mentre juga, és anotar recursos web específics.

5 Generació de coneixement mitjançant jocs online

La idea de fer servir un joc online per generar coneixement, si bé no és nova, no ha estat gaire explotada, i recentment amb els resultats obtinguts per l'investigador Luis von Ahn, sembla que és un camp al qual se li pot treure molt rendiment

5.1 Breu ressenya històrica

La utilització de jocs online per la generació de coneixement és un fenomen que no és nou. Una mostra de que el concepte no és nou ens el dona 20Q (www.20q.net), una aplicació d'ús comercial que té els seus orígens a l'any 1988. El joc consisteix en que el jugador pensa en alguna cosa (animal, vegetal, mineral, alguna altre cosa) i 20Q en base a un conjunt de 20 preguntes que el jugador ha de respondre, endevina que és el que aquest ha pensat. 20Q implementa un algorisme de AI que per cada partida jugada, adquireix més coneixement.

5.2 Human Computation

L'investigador Luis von Ahn, en treballs recents, ha portat el concepte de 'Human Computation' a la categoria de paradigma per solucionar problemes computacionals complexes que fins ara no es sabia com afrontar.

Hi ha gran quantitat de tasques que les persones saben fer fàcilment i que en canvi els ordinadors són incapaços de fer. Malgrat els grans avenços tant en el camp de la informàtica com en el de la intel·ligència artificial, una tasca com és el reconeixement de imatges, és trivial de fer per les persones, però molt complicat per un ordinador. En aquest sentit, podem considerar a cada persona com una unitat de processament extremadament sofisticada, la qual, si tenim en compte a tota la humanitat, s'integra en un sistema de processament distribuït i de gran escala, que anomenaríem "sistema de computació humana".

Les persones però, a diferència dels ordinadors, necessiten un incentiu si volem que facin determinades tasques. Aquest incentiu podria ser econòmic, però la magnitud de determinats problemes impliquen o bé que s'ha de fer una gran inversió, o bé que s'ha de dedicar molt de temps a solucionar el problema, amb la qual cosa també s'acaba convertint en una inversió inviable.

Segons von Ahn, l'any 2003, a tot el món, es van jugar 9 bilions d'hores al famós joc de PC del Solitari¹. Donada aquesta dada, la proposta de von Ahn és dissenyar jocs online (plataforma web), els quals, de manera subjacent, plantegin problemes complexes i aprofitar aquestes hores que les persones dediquen voluntàriament a jugar, per solucionar problemes concrets. Són el que von Ahn a anomenat '**Games with a purpose**'².

¹ <http://video.google.com/videoplay?docid=-8246463980976635143>

² <http://www.cs.cmu.edu/~biglou/ieee-gwap.pdf>

5.3 Games with a purpose (GWAP)

L'equip d'investigació de Luis von Ahn ha dissenyat i implementat una sèrie de jocs (part dels quals es troben a <http://www.gwap.com>) seguint el paradigma GWAP, dels quals, seguidament se'n fa la descripció de dos dels més representatius (ESP Game i Verbosity), per veure'n les característiques i els principis de disseny amb que estan basats.

5.3.1 ESP Game

Joc online que es juga per parelles anònimes i sense la possibilitat de comunicar-se entre elles. A cada un dels dos jugadors se'ls presenta un a foto i ambdós jugadors han d'entrar paraules clau (tags) que descriguin la foto que estan veient. Quan hi ha una coincidència en el tag entrat pels dos jugadors, aquests reben punts i passen al següent torn del joc, on se'ls dóna una nova foto i es repeteix el procés. La partida està limitada per un temps màxim de joc, i l'objectiu és aconseguir tants punts com sigui possible.



Figura 1. Interfície d'usuari d'un dels dos jugadors



Figura 2. Seqüència d'un torn del joc. Cap dels dos usuaris pot veure els moviments de la seva parella

Actualment, Google ha comprat una llicència del ESP Game¹ i les metadades que genera el joc s'utilitzen pel seu cercador d'imatges. Aquest cercador, fins abans de la implantació del joc, cercava les imatges en base al nom del fitxer de la imatge i al text on es trobava contextualitzada aquesta imatge. Per exemple, si el criteri de cerca era "casa", el cercador buscava fitxers on el nom d'aquest contingúés el terme casa, o que el text que envoltava la imatge contingúés el terme.

La utilització de ESP Game (que Google ha anomenat Google Image Labeler) ha millorat molt els resultats de les cerques. De fet, ara el cercador és capaç fins i tot de categoritzar. És a dir, distingeix perfectament les consultes "red car" i "blue car".

Per més detalls sobre el joc, es pot consultar l'article científic que es troba a la URL <http://www.cs.cmu.edu/~biglou/ESP.pdf>

¹ <http://images.google.com/imagelabeler/>

5.3.2 Verbosity

Joc online que es juga per parelles anònimes i sense la possibilitat de comunicar-se entre elles. El propòsit del joc és recollir sentències que continguin fets de sentit comú, com per exemple “les persones tenen dues cames”. La partida dura un temps limitat i funciona per torns. A torn alternatiu s’assigna el rol de “Descriptor” a un dels membres de la parella i a l’altre el de “Endevinador”. Al Descriptor se li dona una paraula clau i sobre aquesta ha de construir sentències amb fets de sentit comú. La sentència no pot contenir la paraula clau. Aleshores, aquesta sentència es passa al Endevinador el qual ha d’endevinar de quina paraula clau es tracta.



Figura 3. Interfície d'usuari d'un dels dos jugadors

Per més detalls sobre el joc, es pot consultar l'article científic que es troba a la URL <http://www.cs.cmu.edu/~biglou/Verboosity.pdf>

5.3.3 Característiques de disseny

Hi ha unes característiques bàsiques que compleixen els dos jocs que hem vist en aquest apartat, que són molt importants:

- Es juga per parelles.
- Les parelles es creen de manera aleatòria (els membres no es coneixen).
- Mentre s'està jugant, els membres de la parella no es poden comunicar.
- Una partida té un temps limitat durant el qual es juguen tants torns de joc com és possible.

El disseny del jocs ESP Game o Verboosity, i en general de tots els fets per von Ahn, està pensat des de la perspectiva de disseny d'un algoritme, en el sentit que donada una entrada obtenim una sortida, la qual l'algoritme ha d'assegurar que és correcte. Això, aplicat al ESP Game vol dir que donada una foto d'un cotxe, l'algoritme ha de assegurar que al final del torn de joc aquesta foto ha de quedar correctament associada a una sèrie de paraules claus rellevants per la foto donada. El fet de que el joc es jugui per parelles i aquestes no es coneguin i no es puguin comunicar, és una de les mesures que s'han adoptat per assegurar la correctesa de la sortida de

l'algoritme. Les dues persones han de coincidir en el resultat i no es poden posar d'acord per tal de falsejar-lo.

Una altra característica que compleixen aquests jocs, és que són divertits. Les dades recollides von Ahn així ho demostren, ja que hi ha jugadors que dediquen fins a 40 hores setmanals a jugar. Si bé no hi ha una recepta per aconseguir aquest objectiu, si que hi ha una sèrie de condicions que ajuden a fer el joc atractiu, essent la principal el fet que es jugui per parelles i que els seus membres hagin d'arribar a un acord. Això desperta un sentiment d'unió entre els membres de la parella que es revela molt estimulants per seguir jugant.

Un altre factor important són les llistes de puntuacions, ja que hi ha molt jugadors amb un fort sentit de la competició que es marquen com objectiu sortir els primers de la llista. Per estimular més aquests jugadors, els punts d'un jugador són acumulatius, de manera que la puntuació parcial aconseguida en una partida s'acumula al total, i es sobre aquest que s'elaboren les llistes de jugadors més puntuats. A més, si el jugador està registrat i loginat en el sistema, els seus punts es mantenen entre diferents sessions de joc.

Finalment, cal comentar que hi ha una altra sèrie de característiques que complementen les que s'han explicat, que si bé no són tant rellevants en la conceptualització del joc, si que són de gran importància. Entre d'altres, podríem destacar les següents:

- Capacitat de poder jugar un sol jugador en cas de no poder crear la parella. El jugador juga amb una parella fictícia simulada pel sistema.
- Periòdicament es fan partides control per verificar l'honradesa del jugador (només per aquells que estan loginats).
- Si es detecta la possibilitat que un jugador està jugant de manera deshonest (entrant resultats equivocats a propòsit), no es graven les dades que aquest esta entrant, per la partida actual.

5.3.4 Tipologies de joc

En el transcurs del disseny dels seus jocs, von Ahn ha identificat dues tipologies de jocs les quals ha anomenat "Symmetric verification game" i "Asymmetric verification game", a les quals hi pertanyen respectivament els jocs ESP Game (Symmetric verification game) i Verbosity (Asymmetric verification game) que s'han descrit anteriorment.

Com hem comentat anteriorment, els jocs es caracteritzen per que tenen una entrada (INPUT) i una sortida (OUTPUT), i és sobre el tractament d'aquests senyals que es fa la caracterització de les tipologies.

Symmetric verification game

Els dos jugadors reben el mateix INPUT i cada un d'ells elabora un OUTPUT. Si les dues sortides coincideixen, es considera que l'OUTPUT està verificat i es dona l'algoritme per acabar.

Aquesta tipologia és adequada per quan un INPUT té un nombre de possibles OUTPUTS limitat. D'altre manera seria difícil que els dos jugadors determinessin la mateixa sortida, i el joc no seria divertit.

Asymmetric verification game

Els dos jugadors (P1 i P2) tenen rols diferents en el transcurs d'un torn del joc. Es dóna l'INPUT a P1, el qual elabora una OUTPUT. Aquest OUTPUT és dóna a P2, el qual ha de deduir quin era l'INPUT corresponent a l'OUTPUT que se li ha passat. Si P2 determina correctament l'INPUT, es considera que l'OUTPUT està verificat i es dóna l'algorisme per acabar.

Aquesta tipologia és adequada per quan un OUTPUT té un nombre de possibles INPUTS limitat. D'altre manera seria difícil que P2 donat el OUTPUT pogués determinar-ne l'INPUT, i el joc no seria divertit.

Per acabar, cal destacar que, com hem pogut veure, les dues tipologies tenen en compte els dos factors principals a tenir en compte al fer el disseny del joc: Per una banda, assegurar la correctesa de la sortida i per l'altre que el joc sigui divertit.

5.3.5 Resultats

Els resultats obtinguts per von Ahn són excel·lents¹ i eleven les característiques que conformen els jocs GWAP a paradigma a seguir a l'hora de dissenyar jocs per la generació de coneixement.

5.4 Anotació de recursos web mitjançant jocs online

Recapitulant sobre el que s'ha exposat fins ara en aquest TFC, a l'apartat 'La web semàntica' hem vist la problemàtica associada a la anotació de recursos web, i en aquest apartat hem vist com es pot aplicar el concepte de 'Human Computation' per la resolució de problemes computacionals complexes.

Tenint presents aquestes idees, ara podem formalitzar el que fins ara hem anat perfilant: el problema d'anotació de recursos per la web semàntica, el podem assimilar a un problema computacional complex i per tant, susceptible de ser tractat com un problema de computació humana amb un joc online.

¹ <http://video.google.com/videoplay?docid=-8246463980976635143>

5.5 Exemples de jocs per la generació de coneixement

Actualment a la xarxa podem trobar diverses propostes de jocs destinats a la generació de coneixement. En aquest apartat se'n descriuen uns quants.

5.5.1 Ontogame

Ontogame (<http://www.ontogame.org>) és un projecte de recerca científica que té per objectiu l'aplicació del paradigma 'Games With A Purpose' (GWAP) com a possible solució als problemes de implantació de la web semàntica (creació i alineació d'ontologies i anotació de recursos). Actualment al seu web podem trobar dos jocs OntoPronto i OntoTube destinats respectivament a la anotació d'articles de la Wikipedia i de videos de YouTube.

Els dos jocs compleixen totes les característiques GWAP que s'han descrit anteriorment i que podem resumir amb els següents punts:

- Joc online multiusuari
- Es juga per parelles. Les parelles són aleatòries i no es poden comunicar
- Si un jugador queda desaparellat juga amb una parella fictícia simulada pel sistema.
- Una partida té un temps limitat durant el qual es juguen tants torns de joc com és possible.
- Es fan llistats de puntuacions. Cal que els usuaris estiguin loginats.

La implementació del jocs té les següents característiques

- Aplicació web amb estructura client - servidor, basada en Java.
- S'ha desenvolupat un *framework* que permet la implementació de diferents escenaris de jocs de manera fàcil i ràpida.
- El joc córrer en un servidor Apache Tomcat 5.5.
- Utilitza el repositori RDF Sesame per la persistència de les dades.

OntoPronto (anotació d'articles de Wikipedia)

El joc és de tipologia "Symmetric Verification Game" segons la classificació GWAP. Es presenta al jugadors un article de la Wikipedia i el joc consisteix en respondre una sèrie de preguntes que el sistema fa respecte a l'article. Quan els dos jugadors coincideixen amb la resposta guanyen punts i se'ls fa la següent pregunta. Amb la resposta de cada pregunta el sistema el que fa és anar classificant l'article dins de l'ontologia Proton¹. Així, la primera pregunta demana als jugadors que diguin si l'article parla d'una Classe (concepte genèric) o de una Instància (un objecte concret d'una determinada classe).

OntoTube (anotació de videos de YouTube)

El joc és de tipologia "Symmetric Verification Game" segons la classificació GWAP. El funcionament és semblant al del joc anterior. Es presenta un video als jugadors i aquests han d'arribar al consens respecte a una sèrie de

¹ <http://proton.semanticweb.org/>

preguntes que se'ls fa respecte al video. En aquest cas, els creadors del joc van crear una ontologia específica per descriure videos.

Valoració:

El joc no és divertit degut a que el jugador no pot aportar les seves respostes. Per cada pregunta, les possibles respostes ja estan definides i l'únic que ha de fer el jugador és triar la resposta que millor encaixa a la pregunta feta. El jugador en comptes de divertir-se donant les seves respostes, el que ha de fer és classificar.

5.5.2 OntoPair: Towards a Collaborative Game for Building OWL-Based Ontologies

És un projecte de recerca científica¹ de la National University of Ireland, de Galway, que té per objectiu la creació d'ontologies mitjançant un joc online. El joc compleix totes les característiques del paradigma GWAP, està implementat amb java i és de codi obert. Actualment no hi ha disponible cap versió online per poder-hi jugar.

El joc consta del que els autors han anomenat dues fases, que en realitat són dos jocs diferents. La fase 1 l'han anomenat *Collecting Properties* i la fase 2 *Collecting common sense facts*. Les metadades que s'aconsegueixen a la fase 1 es fan servir com a dades d'entrada de la fase 2.

La mecànica de funcionament de la fase 1 és exactament la mateixa que la del joc ESP Game descrit anteriorment en aquest apartat, amb l'excepció que en comptes d'assignar paraules clau a una foto, els que s'ha de fer és donar paraules clau de les propietats de l'objecte que mostra la foto.

L'objectiu i la mecànica de funcionament de la fase 2 és exactament la mateixa que la del joc Verbosity descrita anteriorment en aquest apartat.

Així doncs, a la fase 1 es recullen propietats d'un objecte determinat, a la fase 2 es recullen sentències de sentit comú d'aquest objecte i les seves propietats, i amb totes aquestes dades, es construeixen ontologies basades en OWL sobre l'objecte en qüestió.

Per exemple, a la fase 1 es mostra la foto d'un llibre, i es recullen les propietats "Títol" i "Autor". A la fase 2 es recullen sentències del tipus "Un llibre té un títol", "Un llibre té com a mínim un autor", i finalment es construeix l'ontologia per llibre.

5.5.3 Listen Game - A game-based approach for collecting semantic annotations of music

És un projecte de recerca científica² de la University of California de San Diego que té per objectiu la recollida de metadades sobre temes musicals, les quals posteriorment s'utilitzen per entrenar un algorisme d'anotació automàtica supervisada.

¹ <http://mklab.iti.gr/CISWeb/proceedings/paper7.pdf>

² http://ismir2007.ismir.net/proceedings/ISMIR2007_p535_turnbull.pdf

El joc té disponible una versió online a l'adreça <http://www.listengame.org/>. La implementació s'ha fet amb un applet de java, i es basa en la idea de 'Human Computation' de von Ahn, per bé que no compleix totes les característiques GWAP descrites anteriorment.

El joc consisteix en que a un grup de jugadors se'ls dóna un fragment musical escollit aleatòriament d'un conjunt de 250, i sis paraules claus escollits també aleatòriament d'un conjunt preestablert de 174. Cada jugador, un cop ha escoltat la música, ha de triar quina de les sis paraules descriuen millor la música i quina la descriu pitjor. En base a les decisions de cada jugador s'elabora una puntuació que reflexa el consens de quines són les paraules que millor i pitjor descriuen al fragment musical, i en base a això es donen els punts a cada jugador.

Una partida consta de vuit tornos de joc. Els set primers segueixen la mecànica descrita i l'últim torn, en comptes d'escollir entre sis paraules preestablertes, el jugador ha de proposar-ne una, la qual es farà servir en partides posteriors, així s'aconsegueix enriquir el vocabulari.

Les especificacions del joc, no aclareixen quants jugadors intervenen en una partida, ni com es sincronitza. Tampoc queda clar si hi ha una sola partida que englobi tots els jugadors o en canvi els jugadors es reparteixen en diferents partides. Això es molt important de cara a recollir informació de qualitat, ja que en el cas que només es jugui una sola partida, diversos jugadors podrien acordar una data i hora per jugar i jugar deshonestament enredant al sistema amb dades falses. En aquest sentit, i seguint les característiques GWAP establertes per von Ahn, la sortida del joc (les dades que es recullen) no estarien verificades i no podríem donar l'algorisme per bo. Per altre banda, les respostes que pot donar el jugador no són obertes, sinó que ha de triar dues paraules entre sis possibles i per passar al següent torn de joc aquests paraules no cal que coincideixin amb les dels altres jugadors, amb la qual cosa es perd el sentiment d'unió entre els jugadors que segons von Ahn és un gran estímul per seguir jugant.

5.5.4 FACTory

FACTory (<http://game.cyc.com/>) és un joc fet amb el propòsit de crear coneixement per un sistema d'intel·ligència artificial anomenat Cyc. Per tant, el propòsit del joc no és contribuir directament al desenvolupament de la web semàntica.

Cyc és una iniciativa privada que es va iniciar l'any 1985 i actualment encara està en desenvolupament. FACTory és una de les fonts de coneixement de Cyc, però tal com s'indica a la seva web, no és l'única:

"New assertions are continually added to the KB [knowledge base] by human knowledge enterers"

Descripció

- Joc online multiusuari.
- Juga un sol jugador.
- És obligat estar loginat per poder jugar.

- Finalitat del joc: El sistema planteja una sèrie de preguntes i l'usuari ha de respondre si aquestes són Cert, Fals, o bé si Desconeix la resposta.
- El jugador rep punts a cada pregunta, per bé que no queda clar quin sistema utilitza per calcular els punts.
- No hi ha límit de temps a l'hora de respondre les preguntes.
- Al finalitzar es mostra un llista de les màximes puntuacions.

Implementació

- La part del client està implementada amb un applet de Java.

5.5.5 Altres exemples

Per no estendre'ns en l'aplicació, direm que hi ha altres iniciatives entre les que podem citar:

1. PhotoSlap: A Multi-player Online Game for Semantic Annotation
<http://www.aaai.org/Library/AAAI/2007/aaai07-215.php>
2. LabelMe
<http://labelme.csail.mit.edu/browser.html>
3. Common Consensus: a web-base game for collecting commonsense goals
<http://www.media.mit.edu/~lieber/Publications/Common-Consensus.pdf>

5.6 Valoració de la situació actual

La situació actual dels jocs online per la generació de metadata semàntica, globalment la podem valorar com a incipient. Aquest fet però no és estrany, si considerem que, per una banda, la mateixa web semàntica es troba també en un estadi incipient respecte a tot el potencial que presenta, i per altre banda, tenim que el paradigma 'Games with a Purpose' també és recent, i tot just comença a mostrar-se com un solució real a problemes computacionals complexes.

Així doncs, la solució d'un problema computacional complex com és la creació i alineació d'ontologies i la anotació de recursos, mitjançant un joc online, és un enfocament nou a l'hora de superar el coll d'ampolla que aquestes tasques representen en el desenvolupament de la web semàntica. Tant és així, que la majoria d'aplicacions que trobem actualment al web, són resultat de projectes de recerca científica.

6 Joc online per l'anotació de blogs

Anteriorment hem exposat que el problema d'anotació de recursos per la web semàntica, el podem assimilar a un problema computacional complex i per tant, susceptible de ser tractat com un problema de computació humana amb un joc online.

Al parlar de recursos web però, estem parlant de manera genèrica, ja que hi ha molta diversitat en les tipologies de recursos disponibles a la xarxa (blogs, llistes de correu, notícies, serveis web, etc). Nosaltres ens centrarem específicament en l'anotació de blogs.

En aquest apartat veurem aspectes generals a tenir en compte a l'hora de fer un joc online per l'anotació de blogs, sense entrar en detall en característiques específiques d'un joc concret.

6.1 Weblogs

Weblog és una paraula resultant dels termes web i log, dels quals finalment s'ha acabat fent la contracció a la paraula **blog** que és el terme que més s'utilitza normalment.

Un blog és una pàgina web que consisteix d'una sèrie d'entrades, que s'anomenen **post**, normalment ordenades per ordre cronològic descendent i habitualment referents a un tema concret que és d'interès de l'autor. Un blog pot tenir un o més autors els quals s'encarreguen de l'edició dels continguts de manera periòdica.

Actualment a la xarxa trobem diferents proveïdors els quals disposen d'eines gratuïtes per que qualsevol usuari pugui crear i mantenir un blog sense necessitat de conèixer les tecnologies web necessàries que el fan possible.

De la definició que acabem de donar, podem veure com efectivament es donen totes les característiques que fan que sigui un recurs difícil d'anotar-ne els continguts per la web semàntica:

- Es poden crear i mantenir sense necessitat de conèixer les tecnologies implicades. Per tant, la gran majoria d'usuaris no disposaran dels coneixements necessaris per fer les anotacions.
- Poden haver-hi diversos autors per un mateix blog. Aquest factor es suma al punt anterior, dificultant encara més la edició de continguts semàntics.
- Els blogs s'editen de manera periòdica. Per tant no es factible la opció d'anotar tots els continguts una primera vegada (feina que podria fer un professional del sector).

6.1.1 Subscripció a blogs

Tots els blogs permeten que els lectors que així ho vulguin, es puguin subscriure a la publicació de continguts, de la mateixa manera que un lector d'un diari de premsa

escrita es pot subscriure al diari i cada dia el rep a casa seva sense necessitat d'anar al quiosc.

Aquest sistema de publicació s'anomena feed, i els blogs fan públics els seus continguts en una URL específica que és l'adreça del feed, a la qual s'hi troba un fitxer XML amb els últims continguts editats pel blog. Actualment es segueixen dos formats a l'hora de publicar aquests fitxers, que són Atom i RSS

Existeix programari que facilita la lectura d'aquests fitxers XML, a la vegada que regularment consulta els feed per comprovar si han canviat els continguts, de manera que no cal haver d'anar a la pàgina del blog per veure si hi ha continguts nous.

6.2 Anotació de blogs

En aquest apartat veurem que donat un post d'un blog, hi ha moltes dades susceptibles de ser anotades. Part d'aquestes dades es poden anotar automàticament, però la part més important que és la que fa referència al contingut dels posts no es pot.

El fet que el feed vingui donat amb un format preestablert (Atom o RSS), ens permet la possibilitat de fer una anotació automàtica, de fet, el fitxer XML del feed ja incorpora metadades. El següent exemple presenta un arxiu XML d'un feed amb format Atom:

```
<?xml version='1.0' encoding='UTF-8'?>
<?xml-stylesheet href="http://www.blogger.com/styles/atom.css" type="text/css"?>
<feed xmlns='http://www.w3.org/2005/Atom' xmlns:openSearch='http://a9.com/[...]'>
  <id>tag:blogger.com,1999:blog-7986754043868005909</id>
  <updated>2008-12-29T09:44:42.918-08:00</updated>
  <title type='text'>Selecció poètica</title>
  <subtitle type='html'>...poemes de poetes poètics.</subtitle>
  <author>
    <name>llampdellamp</name>
    <uri>http://www.blogger.com/profile/00639620987681733743</uri>
    <email>noreply@blogger.com</email>
  </author>
  <generator version='7.00' uri='http://www.blogger.com'>Blogger</generator>
  <openSearch:totalResults>12</openSearch:totalResults>
  <openSearch:startIndex>1</openSearch:startIndex>
  <openSearch:itemsPerPage>25</openSearch:itemsPerPage>
  <entry>
    <id>tag:blogger.com,1999:blog-7986754043868005909.post-10284968450438942</id>
    <published>2008-12-29T09:42:00.000-08:00</published>
    <updated>2008-12-29T09:44:42.935-08:00</updated>
    <category scheme='http://www.blogger.com/atom/ns#' term='Miquel Martí i Pol' />
    <title type='text'>Tanta llum de mar</title>
    <content type='html'>[...]Aquí va el contingut del post[...]</content>
    <author>
      <name>llampdellamp</name>
      <uri>http://www.blogger.com/profile/00639620987681733743</uri>
      <email>noreply@blogger.com</email>
    </author>
    <thr:total xmlns:thr='http://purl.org/syndication/thread/1.0'>0</thr:total>
  </entry>
</feed>
```

Nota: L'arxiu no és complert, i de les dades que es presenten s'han eliminat parts que dificultaven la comprensió.

En aquest arxiu podem veure dues parts. La primera conté dades genèriques del blog (autor, títol, data última edició), i la segona conté dades d'una entrada (post) concreta del blog (autor, títol, contingut, etc). Veiem com totes les dades van

embolcallades per les seves metadades i per tant doncs, ja sabem la semàntica dels continguts del fitxer.

De tots els camps que té el fitxer, n'hi ha un que per nosaltres és d'especial interès, que és el camp `content`. En aquest camp, i tal com el seu nom indica, hi trobem el contingut del post, i el que és important de veure és que d'aquest si que no en sabem quina és la semàntica. Senzillament és un bloc de text que sabem que és el contingut, però no sabem de que parla.

El camp `category`, acompanya el de `content`, i és un camp que permet a l'autor posar paraules claus que descriguin el camp `content`. Per tant, aporten un cert nivell semàntic al contingut, però és de molt baix nivell i a efectes pràctics només serveix per classificar el post en categories. Cal destacar que hem dit que aquest camp, "només" serveix per classificar, però el ventall de possibilitats que obre és molt gran, ja que és la clau que facilita que moltes aplicacions puguin treballar, com per exemple el cercador de blogs `technorati.com`. Només d'aquest fet, podem imaginar la rellevància que podria tenir el fet que els continguts estiguessin degudament anotats.

La solució que proposa aquest TFC és dissenyar jocs online que permetin fer l'anotació d'aquests continguts.

6.2.1 Anotació automàtica

En referència a la anotació automàtica que podem fer d'un feed, podem citar que hi ha iniciatives en aquest sentit. Com a exemple, a la web de SIOC, s'hi troba una eina¹ que permet anotar amb el vocabulari SIOC posts generats amb wordpress (eina d'edició de blogs).

6.3 Criteris de disseny

A l'hora de fer el disseny d'un joc per l'anotació de blogs, hi ha una sèrie de criteris generals a tenir en compte independentment de quin sigui el joc concret que és dissenyi. El més important és que el joc sigui divertit, i en aquest sentit cal tenir presents els criteris del paradigma GWAP, però també cal tenir en compte la temàtica del blog.

Un blog acostuma a tractar algun tema concret i per tant els seus posts acostumen a fer referència a diferents aspectes de la temàtica central del blog. A l'hora de pensar en la mecànica del joc, podem tenir en compte la temàtica i fer jocs específics, o en canvi podem pensar en un joc que no tingui en compte la temàtica, i que per tant ens permeti anotar tota mena de blogs.

Pel cas de jocs genèrics que no tinguin en compte la temàtica, resulta difícil trobar jocs divertits per tot els ventall possible de jugadors. Donat que a través del joc volem que el jugador ens expliqui quin és el significat del contingut dels posts, una activitat central del joc consistirà en la lectura d'aquests posts. Si el joc no té en

¹ <http://sioc-project.org/wordpress>

compte la temàtica, el jugador es trobarà amb posts de contingut heterogeni i possiblement part d'ells no els trobarà interessants, amb la qual cosa correm el risc de que deixi de jugar.

Pel prototip que s'ha desenvolupat en aquest TFC, s'ha tingut en compte aquest factor, i s'ha dissenyat un joc tenint en compte la temàtica. La temàtica escollida ha estat la poesia.

Un altre factor a tenir en compte, és com accedirem a les dades per tal de presentar-les als jugadors. En aquest cas farem ús del sistema de subscripcions (feeds) els quals ens proporcionen un fitxer XML amb els continguts del blog. Sobre els feeds però, hem de tenir en compte els següents punts:

- El fitxer XML no conté tots els post del blog, sinó que només conté un determinat nombre dels posts més recents. Això vol dir que si volem accedir a un post antic, potser que aquest ja no el trobem. Si l'activitat del blog és molt alta, hi ha posts que en l'interval d'un dia ja no s'inclouen al fitxer. De tota manera, és possible descarregar d'un blog posts que ja no estan inclosos en el feed.
- L'autor d'un blog pot editar els posts en qualsevol moment. Això vol dir que es pot crear un post i posteriorment modificar el seu contingut, Si bé aquesta situació no és freqüent, si que és possible. Això caldrà tenir-ho en compte en el cas que per l'arquitectura del joc, s'hagi decidit guardar el contingut dels post en una base de dades local.
- Hi ha feeds que no proporcionen el contingut dels posts, sinó només una breu descripció. Això vol dir que no ens serà possible presentar el contingut del post mitjançant la interfície gràfica del nostre joc. Aleshores tenim dues alternatives: la primera és seguir l'enllaç que sempre es proporciona amb cada entrada del feed i visitar el web on es troba el blog, però ens arrisquem a que el jugador perdi l'atenció sobre el joc. L'altre opció seria encastar tota la pàgina del web original a dins de la interfície gràfica del joc.

Finalment, un altre factor a tenir en compte, és l'elecció de les ontologies adequades que millor ens permetin anotar els continguts. Per aconseguir això, cal conèixer quines són les ontologies desenvolupades per tercers que més ens convenen. Sinó l'alternativa és crear una ontologia pròpia.

6.4 Ontologies útils per l'anotació de blogs

Una ontologia molt indicada per l'anotació de blogs, és l'ontologia SIOC¹ la qual està pensada per l'anotació de recursos web generats per comunitats online, i que per tant s'ajusta molt al nostre propòsit.

Una de les característiques de SIOC, és que des de l'inici està concebuda amb la intenció de reutilitzar altres vocabularis ja existents, evitant així redundància en les definicions, a la vegada que permet una definició més rica en dominis de coneixement específics. En aquest sentit, és habitual l'ús de diverses ontologies juntament amb SIOC, com són:

- Dublin Core: <http://dublincore.org/>
- FOAF: <http://xmlns.com/foaf/spec/>
- RSS 1.0: <http://web.resource.org/rss/1.0/>
- SKOS: <http://www.w3.org/2004/02/skos/>

¹ <http://www.w3.org/Submission/sioc-spec/>

7 Prototip – Requisites

Hem vist les característiques generals que hauria de tenir un joc per l'anotació de blogs. Ara les posarem en pràctica centrant-nos en una aplicació concreta, i en veurem els detalls.

7.1 Finalitat del joc

El prototip que s'ha desenvolupat l'hem anomenat **Angam**, i és un joc per l'anotació de blogs de temàtica poètica. El joc té dos objectius principals. En base a un post que es presenta al jugador, el primer objectiu és determinar si el post conté o no una poesia i en cas afirmatiu, el segon objectiu és descriure què és el que expressa la poesia en qüestió.

7.2 Criteris de disseny

A l'hora d'afrontar el disseny, hem intentat cenyir-nos als criteris de disseny dels 'Games With A Purpose' i hem volgut complir la característica principal, que és que el joc es juga per parelles.

Per altre banda, s'ha decidit que els posts que s'utilitzaran a les partides, els tinguem en una base de dades relacional al nostre servidor. Aquesta decisió bé motivada pel fet que, donat que el joc té la poesia per temàtica central, hem considerat que el volum de posts que tindrem no serà molt gran, si més no comparat amb el volum que suposaria un joc genèric. A més d'aquesta manera també accelerarem el temps de resposta del joc, ja que no caldrà anar a buscar el post a la web del blog, sinó que el servirem directament de la nostra base de dades. Al tenir un joc ràpid, també aconseguirem mantenir l'atenció del jugador, és més dinàmic.

Per fer encara més atractiu i dinàmic el joc, s'ha decidit que les interaccions client – servidor es faran utilitzant tècniques Ajax, amb l'objectiu de que al no haver de recarregar la pantalla a cada comunicació client – servidor, mantindrem més l'atenció del jugador.

Finalment, només ens falta decidir quines ontologies farem servir per fer la anotació dels nostres posts. Utilitzarem SIOC com ontologia principal. Per altre banda, volem descriure què és el que expressa una poesia, i aquí el ventall és molt ampli. Pot ser un sentiment, un objecte, una persona, etc. Generalitzant, podem dir que en són Conceptes. La ontologia SKOS, disposa de la classe Concept, que s'ajusta perfectament al nostre propòsit.

7.3 Funcionament del joc

El joc consta de cinc torns de joc. A cada torn es presenta un mateix post als dos jugadors. Els jugadors han de determinar si el post conté algun poema i si el conté, han d'entrar paraules claus que el descriguin. Quan el sistema troba una coincidència a les paraules clau del dos jugadors, els premia amb punts, i automàticament es passa al següent torn de joc.

El joc disposa d'un botó 'No té contingut poètic' que serveix als jugadors per indicar que el post actual no conté cap poesia. Si els dos jugadors arriben a un acord en aquest sentit, es passa al següent torn. També es disposa d'un botó 'Següent torn' per si es vol passar al següent post sense haver arribat a un acord amb l'altre jugador.

7.4 Nota d'aclariment

Com veurem, l'anàlisi de l'aplicació s'ha fet en base al requisit de que el joc és jugat per parelles. Finalment aquesta funcionalitat no s'ha implementat i a l'etapa de disseny es justificarà el perquè.

Per tant, cal aclarir doncs, que la implementació del joc és per un sol jugador, però l'anàlisi s'ha fet tenint present el joc per parelles. De la mateixa manera, a la implementació final, els jugadors estan obligats a loginar-se en el sistema per poder jugar, i l'anàlisi s'ha fet tenint en compte que podíem tenir jugadors anònims (no registrats).

8 Prototip – Anàlisi

8.1 Modes de Joc

El joc és per parelles, però al iniciar una partida es pot donar el cas que el número de jugadors disponibles sigui imparell, amb la qual cosa un jugador queda desaparellat.

Es pot donar el cas també, que en una partida en curs, algun dels dos jugadors abandona, amb la qual cosa la partida quedaria inacabada.

Per cobrir aquests cassos, cal tenir dos modes de joc diferents: Two-Players i Single-Player:

8.1.1 Mode de joc Two-Players

És el mode de joc predefinit per defecte. Quan s'inicia una partida el sistema intenta crear parelles.

8.1.2 Mode de joc Single-Player

Si a l'hora d'iniciar una partida no es pot crear la parella, o bé en una partida en curs en mode Two-Players un jugador abandona, s'ha d'activar el mode Single-Player, on el sistema ha d'assumir el rol de segon jugador, produint l'efecte sobre l'altre jugador de que està en una partida normal.

8.2 Actors

Seguidament veurem els actors que intervenen en el sistema i les seves funcions.

8.2.1 Visitant

Un Visitant és qualsevol persona que accedeix al web on s'hostatja el joc. Tot Visitant que accedeixi al web pot jugar de forma anònima, és a dir sense identificar-se al sistema. Quan un usuari s'identifica, es converteix en **Usuari**.

8.2.2Usuari

És un Visitant que s'ha registrat amb usuari i password en el sistema i aquest l'ha reconegut positivament.

8.2.3 Jugador

Un jugador és un Visitant o unUsuari el qual ha iniciat una partida. Si inicialment era unUsuari les accions que faci durant les partides que jugui, seran registrades pel **Sistema**. Això permetrà fer estudis de les dades recollides, per usuari. Si en canvi el jugador inicialment era un Visitant (no registrat al sistema), el jugador serà anònim, i les dades que generi no ponderaran tant.

Sobre els usuaris registrats, es poden fer partides de control per comprovar la honestedat de les dades que generen. Això no es pot fer amb jugadors anònims i per tant les seves dades han de tenir un pes menor ja que no es pot controlar la qualitat de les dades que generen.

El jugador pot abandonar la partida en qualsevol moment.

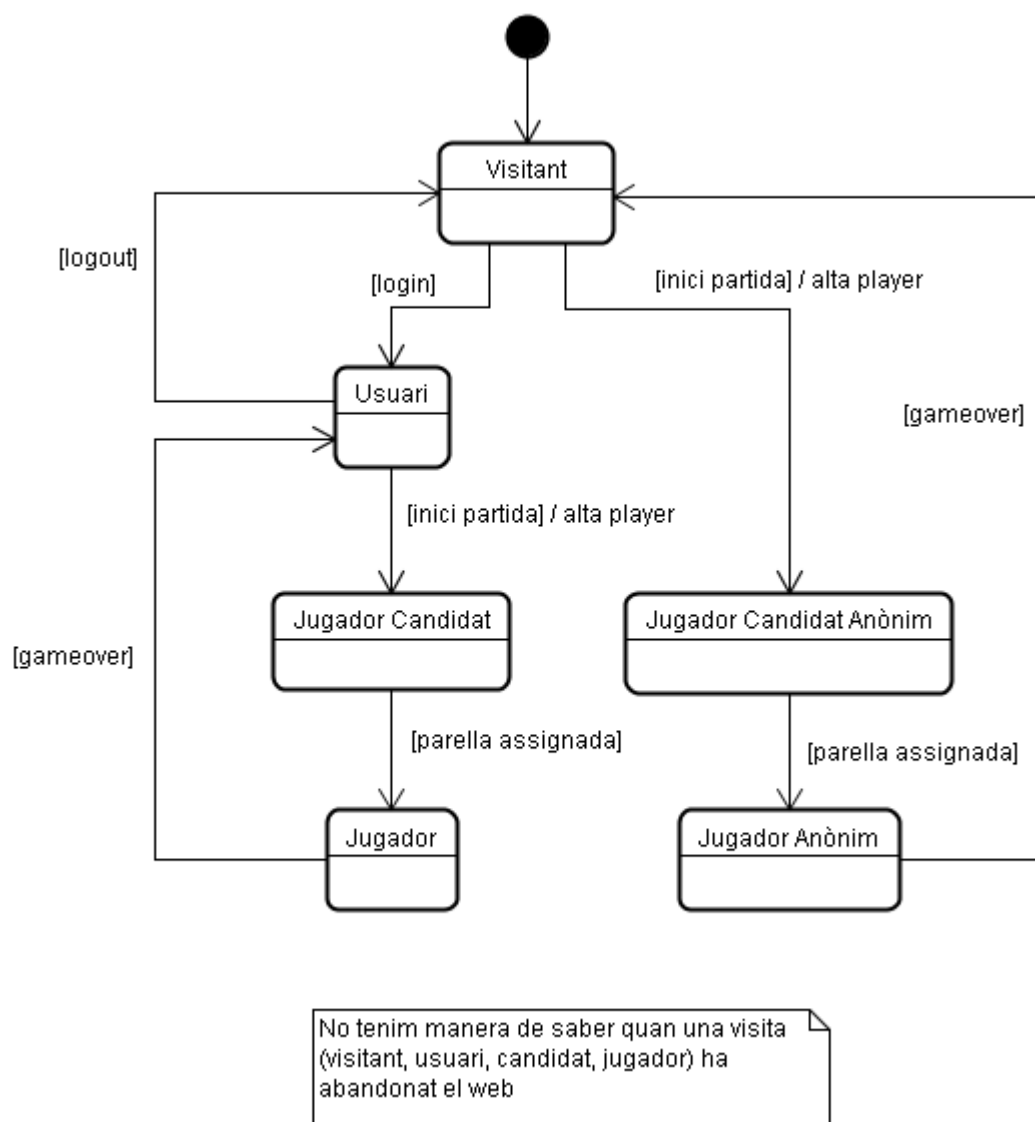
8.2.4 Sistema

El sistema és representat pel servidor on s'hostatja el joc, el qual executa tots els processos que controlen el funcionament del joc. A més, si durant una partida un dels dos jugadors abandona, el sistema ha d'assumir el seu rol i continuar la partida.

8.2.5 Administrador

És la persona que administra el Sistema.

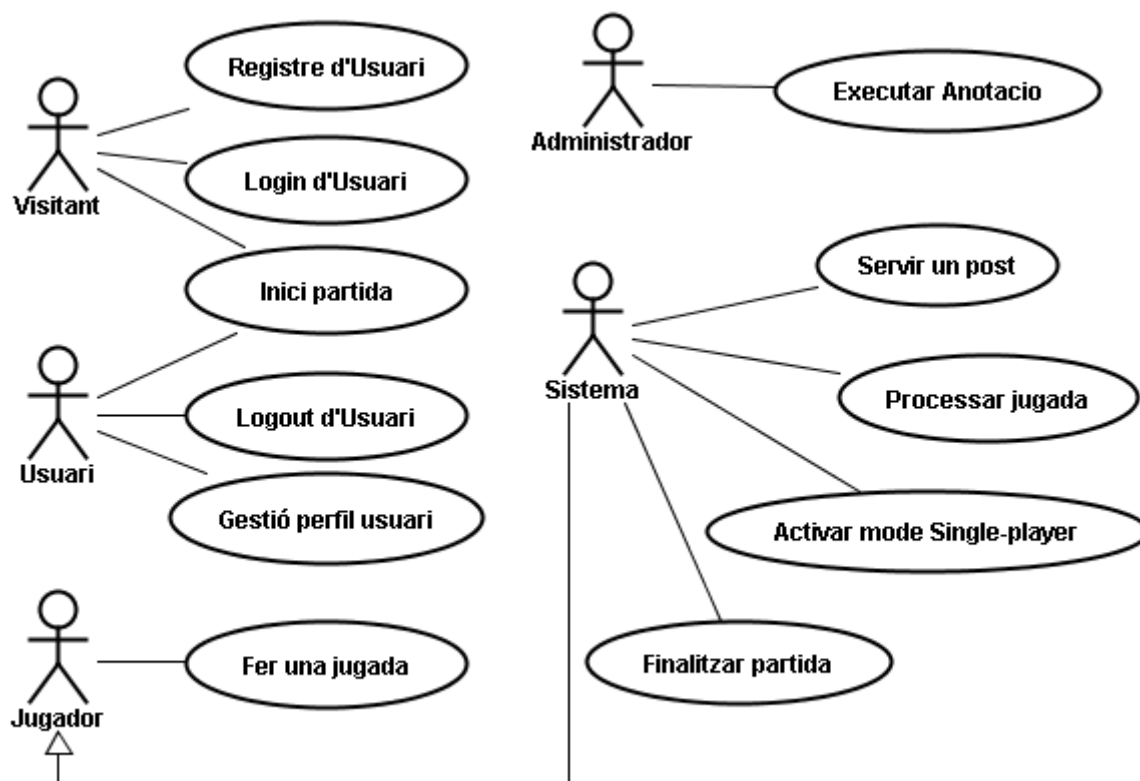
8.2.6 Diagrama d'estats pels visitants del web



Quan un Visitant o un Usuari inicien una partida, passen a ser Jugador Candidat, i es mantenen en aquest estat mentre el sistema els busca una parella per poder jugar. Si es troba la parella la partida arrancarà en mode Two-Players. En canvi, si no existeix cap altre Usuari o Visitant per fer la parella, el sistema assumeix el rol i assigna una parella fictícia. Aquest és el que anomenarem el mode Single-Player.

8.3 Casos d'ús

8.3.1 Diagrama de casos d'ús



8.3.2 Registre d'Usuari

Funcionalitat: Dóna d'alta un usuari al sistema.

Actors: Visitant

Casos d'ús relacionats: Login Usuari

Precondició:

Postcondició: El Visitant ha estat donat d'alta a la base de dades del sistema.

Procés:

1. El sistema demana al Visitant un nom d'usuari i password. El password es demana dues vegades per tal de verificar que és correcte.
2. El sistema demana al Visitant la url d'un *feed* (que suposarem que és el seu). Aquesta opció no és obligatòria.
3. El sistema verifica que els dos passwords són coincidents, i dóna d'alta el Jugador.
4. El sistema verifica que la url del *feed* és vàlida.

Alternatives de procés:

1. Si no es subministren les dades obligatòries o els passwords no són coincidents, no es dóna d'alta al Visitant i se'l informa.
2. Si la url del *feed* no és vàlida, aquesta no s'incorpora al sistema i s'informa al a l'Usuari, donant-li l'oportunitat de rectificar-la o desestimar-la.

8.3.3 Login Usuari

Funcionalitat: Identifica un visitant que vol accedir al sistema.

Actors: Visitant

Casos d'ús relacionats: Registre Usuari

Precondició: El Visitant existeix a la base de dades del sistema.

Postcondició: El Visitant s'ha loginat. Pot iniciar partida.

Procés:

1. El Visitant entra el seu nom d'usuari i password.
2. El sistema comprova si les dades són correctes i logina el Visitant.

Alternatives de procés:

1. Si les dades subministrades no són correctes, no es logina el Visitant i se'l informa.

8.3.4 Gestió del perfil d'Usuari

Funcionalitat: Permet modificar les dades del Jugador.

Actors: Jugador

Casos d'ús relacionats: Login Usuari

Precondició: El Jugador s'ha loginat correctament al sistema.

Postcondició: El Jugador ha modificat correctament les seves dades a la base de dades del sistema.

Procés:

1. El sistema mostra les dades del Jugador.
2. El jugador modifica les seves dades.

Alternatives de procés:

1. Si no es subministren les dades obligatòries o els passwords no són coincidents, no es modifiquen les dades del Jugador i se'l informa.

8.3.5 Inici Partida

Funcionalitat: Desencadena tots els mecanismes per iniciar una partida.

Actors: Visitant, Usuari

Casos d'ús relacionats: Servir un 'post', Activar mode Single-player

Precondició: Un Visitant o un Usuari ha indicat que vol iniciar una partida.

Postcondició: S'ha iniciat una partida en el mode de joc més adequat.

Procés:

1. El sistema busca una parella pel jugador i la troba. S'inicia partida *two-player mode*
2. El sistema busca el *post* més adequat per ser anotar durant la partida, i el serveix als dos jugadors.

Alternatives de procés:

1. El sistema busca una parella pel jugador i NO la troba. S'inicia partida *single-player mode*, en la qual el sistema recupera del log de partides una partida feta anteriorment, i els *posts* d'aquesta partida és la que es serveix al jugador.

8.3.6 Fer una jugada

Funcionalitat: Un jugador fa una jugada i la comunica al sistema.

Actors: Jugador

Casos d'ús relacionats: Processar una jugada

Precondició: S'ha iniciat una partida i aquesta està en marxa.

Postcondició: La jugada ha estat comunicada al sistema.

Procés:

1. Mitjançant la interfície gràfica el jugador comunica al sistema la seva jugada.

Alternatives de procés:

1. Si en el moment de fer la jugada no hi ha comunicació, es reintenta indefinidament.

8.3.7 Processar una jugada

Funcionalitat: Processa una jugada feta per un dels jugadors.

Actors: Sistema

Casos d'ús relacionats: Fer una jugada, Servir un 'post', Avortar partida

Precondició: Un jugador ha fet una jugada.

Postcondició:

Procés:

1. Rep les dades enviades des del navegador web del jugador.
2. Guardar la jugada a la llista de respostes donades pel *post* en curs.
3. Tant si és *Single-player mode* com *Two-player mode*, es busca si la jugada té alguna coincidència amb la llista de respostes del segon jugador:
 - 3.1. Hi ha coincidència,
 - 3.1.1. S'anota la jugada
 - 3.1.2. Es serveix un nou *post*
 - 3.2. No hi ha coincidència: No es fa res.

Alternatives de procés:

1. El sistema al rebre la jugada detecta que l'altre jugador ha abandonat. Aleshores s'avorta la partida.

Pendent d'estudi: Si un jugador abandona el sistema podria adoptar el rol del jugador (activar el *Single-player mode*).

8.3.8 Servir un 'post'

Funcionalitat: El sistema selecciona una entrada d'un blog (*post*) i el presenta als jugadors

Actors: Sistema

Casos d'ús relacionats: Inici partida, Processar una jugada

Precondició: S'ha iniciat una partida i aquesta està en marxa.

Postcondició:

Procés:

1. Es selecciona un blog de la base de dades. La selecció es farà en base als punts acumulats que els propietaris dels blogs hagin aconseguit a les seves partides. Com més punts tingui, més probabilitat tindran que el seu blog sigui el seleccionat per la següent partida. Un jugador no podrà anotar el seu blog.
2. Un cop seleccionat el blog, es seleccionarà l'entrada (*post*). En una mateixa partida no es podran repetir posts.
3. Es presentarà el *post* als jugadors.

Alternatives de procés:

1. Si el sistema està en *Single-player mode*, es seguiran els mateixos criteris de selecció, però a més el *post* que es serveix ha d'haver estat utilitzat en alguna altra partida, de manera que es trobi en el log de partides jugades.

8.3.9 Activar mode Single-player

Funcionalitat: Quan un jugador està sense parella, el sistema assumeix el rol de jugador.

Actors: Sistema

Casos d'ús relacionats: Inici partida

Precondició: A l'iniciar la partida no s'ha trobat parella pel jugador, o bé a mitja partida el sistema detecta que hi ha un jugador que ha abandonat.

Postcondició: El sistema assumeix el rol de jugador.

Procés:

1. S'activa el mode *Single-player*

Alternatives de procés:

8.3.10 Finalitzar la partida.

Funcionalitat: Actualitza els marcadors dels jugadors.

Actors: Sistema

Casos d'ús relacionats:

Precondició: La partida ha finalitzat.

Postcondició:

Procés:

1. Si algun dels dos jugadors estava *loginat*, se li sumen els punts.
2. Es mostra als jugadors l'històric de màximes puntuacions, i es dóna l'opció de tornar a jugar

Alternatives de procés:

8.3.11 Executar Anotació

Funcionalitat: Amb els resultats obtinguts de les partides jugades, extreu metadades sobre els *posts* dels diferents blogs registrats en el sistema.

Actors: Administrador

Casos d'ús relacionats:

Precondició:

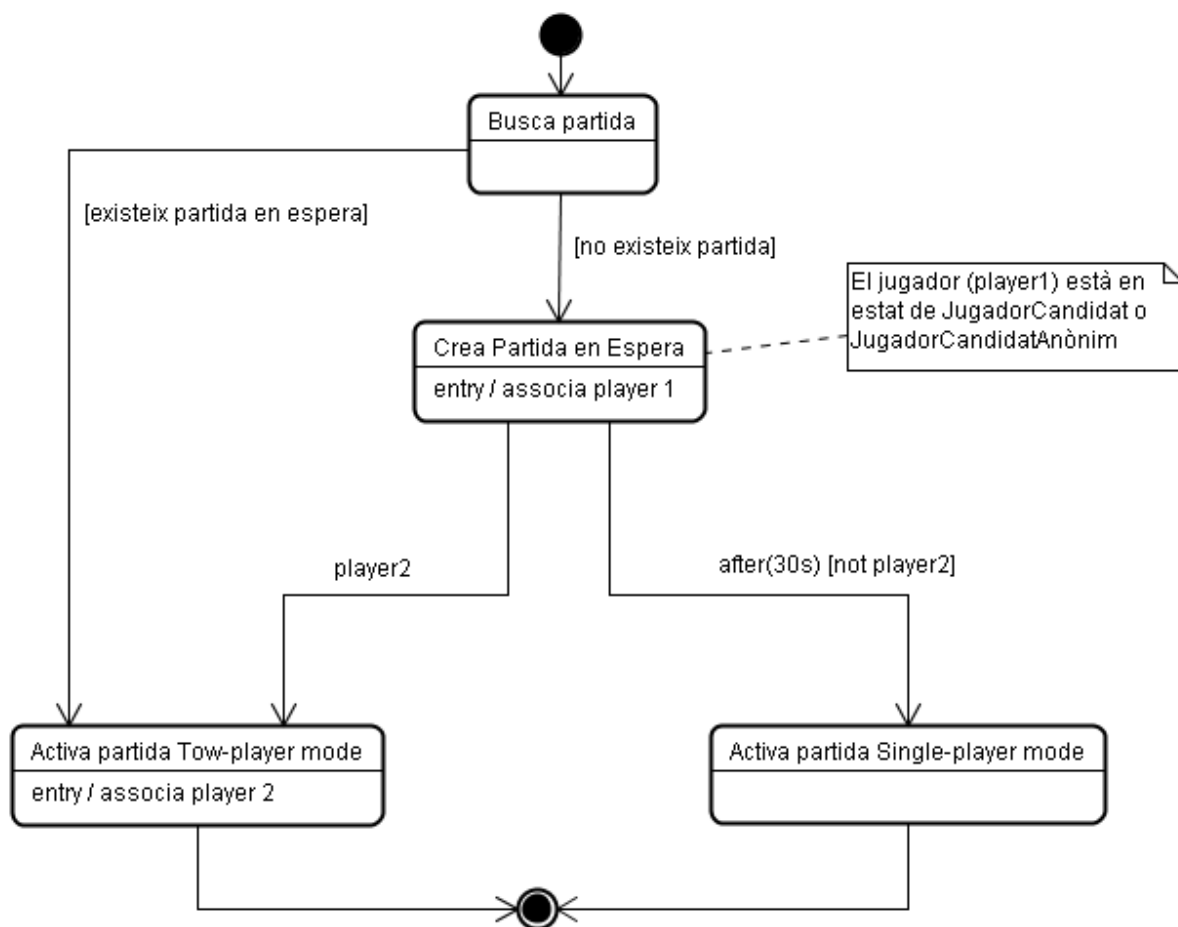
Postcondició:

Procés:

1. Sota demanda, l'administrador del sistema executa el procés que extreu metadata de les partides jugades. Com a resultat s'obté un fitxer RDF.

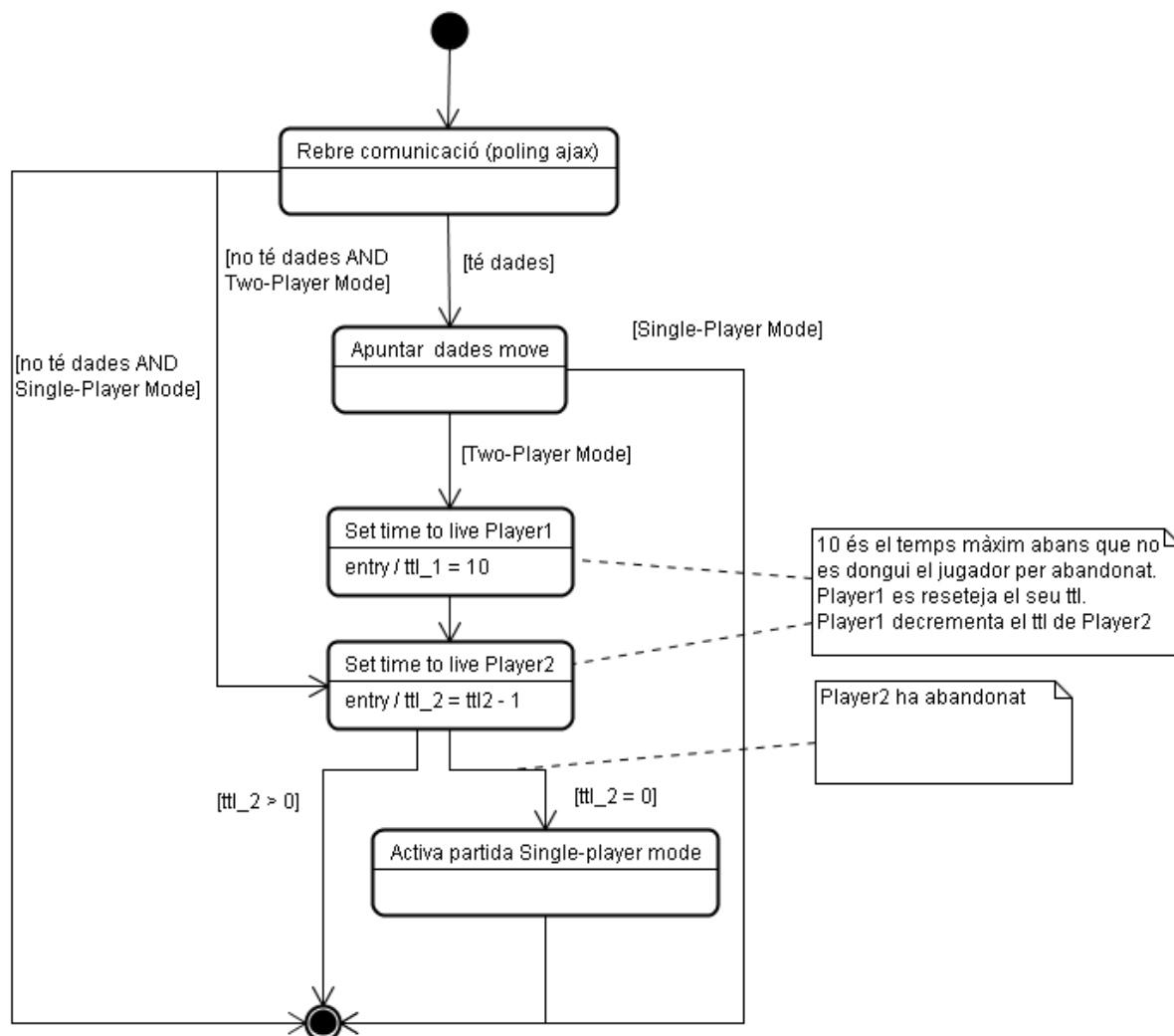
Alternatives de procés:

8.4 Algorisme d'inici de partida



Quan un jugador inicia una partida, el sistema intenta trobar-li una parella. Primer mira si hi ha algun jugador amb partida iniciada i a l'espera de trobar parella. Si el troba assigna el nou jugador i acaba. Si no hi ha cap partida en espera, crea una nova partida i queda a l'espera de que arribi un nou jugador. Si arriba, inicia la partida normalment, sinó, passats uns segons arranca el mode de joc Single-Player.

8.5 Algorisme de detecció d'abandonament d'un jugador



Estant el joc en mode Two-Players, cal controlar que l'altre jugador no abandoni la partida. En cas de fer-ho s'ha d'activar el mode Single-Player.

Escenaris possibles:

- Dos jugadors: Un abandona
- Dos jugadors: Un queda parat durant un temps indefinit, però no abandona i finalment continua jugant. En aquest cas els dos acaben jugant en mode Single-Player.
- Dos jugadors: Els dos queden parats durant un temps indefinit, però cap dels dos abandona i finalment continuen jugant. En aquest cas els dos acaben jugant en mode Single-Player.

9 Prototip – Disseny

9.1 Variacions respecte l'etapa d'anàlisi

A l'etapa d'anàlisi s'havia previst que el joc fos per parelles, i també que hi poguessin haver jugadors anònims (no registrats). Malauradament després de l'etapa d'anàlisi es va veure que no es disposaria de temps suficient per implementar aquestes funcionalitats.

Respecte a la primera funcionalitat (joc per parelles), el fet que el joc sigui mono usuari simplifica molt el disseny:

- No hi ha dos modes de joc.
- Al iniciar una partida, no cal formar parelles.
- No s'ha de controlar si hi ha abandonament d'algun jugador.
- El sistema no ha d'assumir el rol de jugador en cas d'abandonament.

Com veiem el nivell de complexitat del joc per parelles es substancialment més elevat que el de un sol jugador. Això ha fet que per falta de temps, aquestes funcionalitats no s'hagin implementat.

La complexitat de la segona funcionalitat no és tant complicada com la primera, però per falta de temps, tampoc no s'ha implementat.

Malgrat tot el disseny si que “apunta” la possibilitat de que en un futur es pogués desenvolupar aquestes funcionalitats. Per exemple, com veurem, la classe Game té dues referència a la classe Player, en previsió del joc per parelles, així doncs no ens ha d'estranyar que al codi trobem referències a aquestes funcionalitats.

9.2 Funcionament del joc

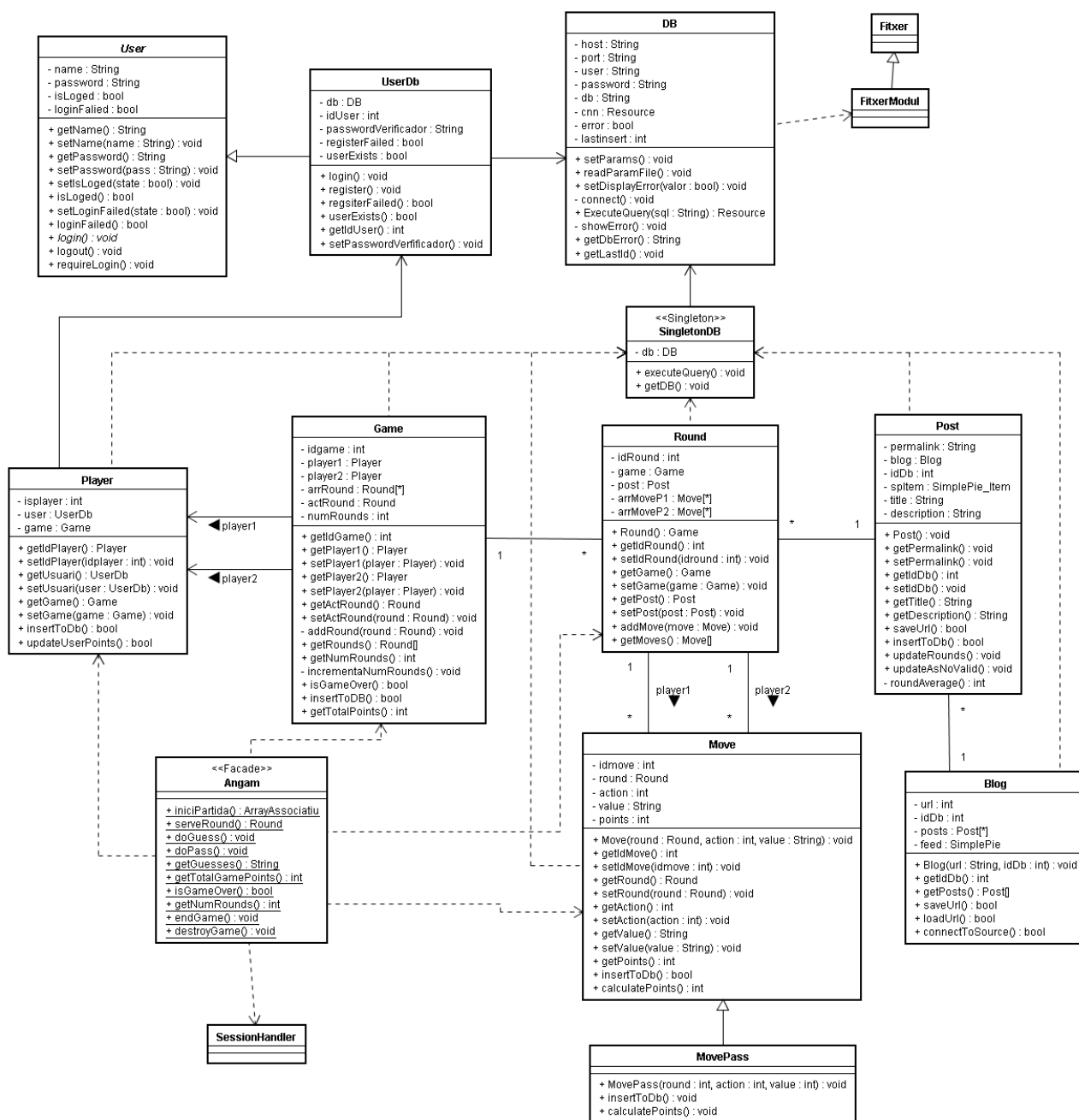
En base a les modificacions comentades, donem una nova especificació del funcionament del joc.

El joc és mono usuari i consta de cinc torns de joc. A cada torn es presenta un post. Per descriure el post, es poden entrar tantes paraules clau com es vulgui i per cada paraula clau el jugador rebrà punts en funció de quants altres jugadors hagin fet la mateixa descripció.

El jugador pot estar el temps que vulgui a cada torn de joc i la partida no té un temps límit. Donat que la temàtica és la poesia i es tracta de extreure el significat d'un poema, s'ha considerat que no s'ha de donar pressa als jugadors.

El joc disposa d'un botó 'Següent torn' que li serveix al jugador per passar al següent post un cop decideixi que ja ha acabat la descripció actual. Hi ha també un botó 'No té contingut poètic' que li serveix al jugador per indicar que el post actual no conté cap poesia. Si es fa servir aquest botó és passa automàticament al següent torn de joc.

9.3 Diagrama de classes



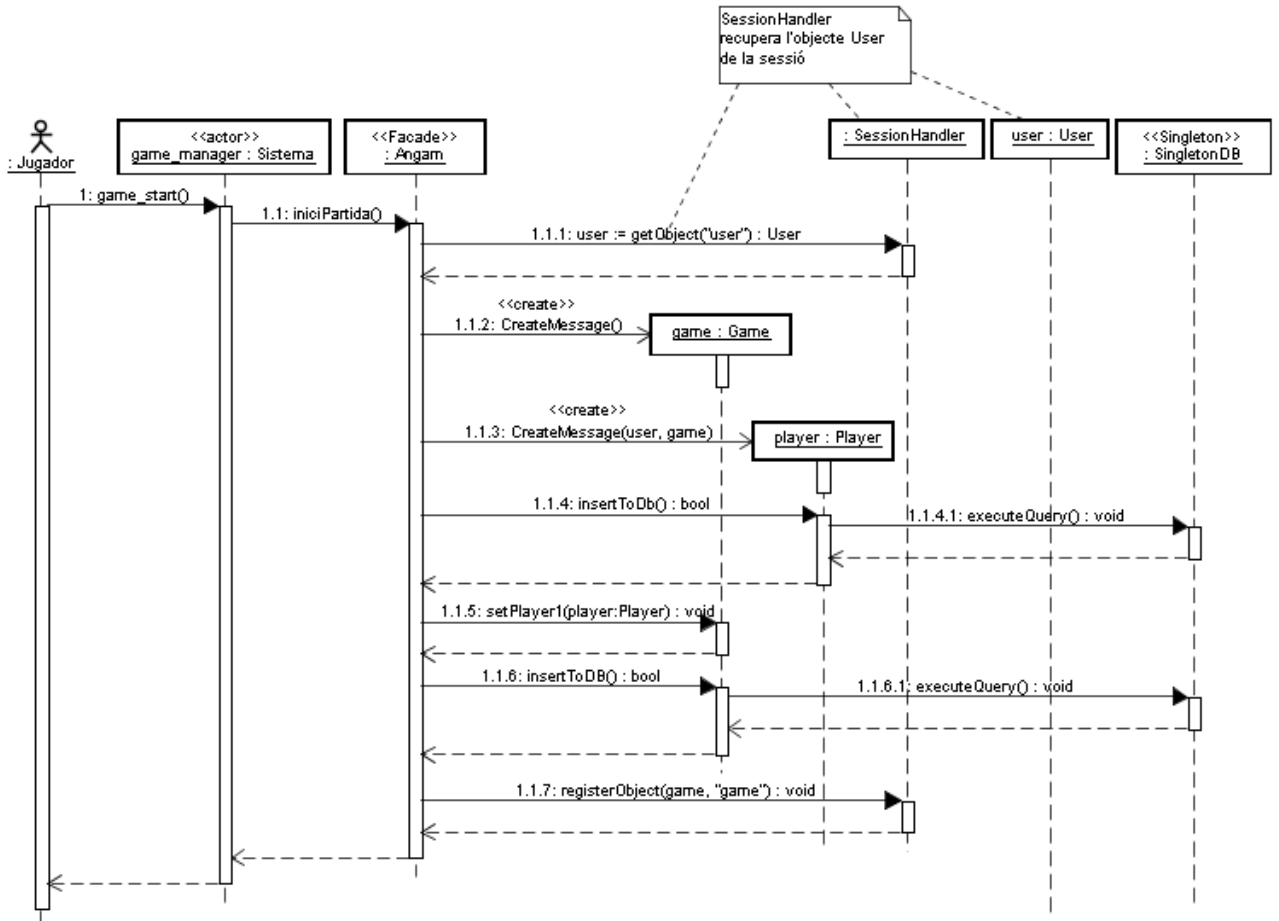
L'essència del disseny és la següent: Un Game està format per un conjunt de Round (torns de joc). Cada Round té assignat un Post, que es el que es serveix als jugadors per que anotin. Cada Round té un conjunt de Move que són les anotacions que els jugadors fan pel post associat al Round.

Aquestes classes que hem comentat són el que podríem anomenar el framework del joc. Per fer ús d'aquest framework s'ha dissenyat la classe Angam la qual implementa el patró de disseny Facade, de manera que una aplicació client pot interactuar amb el framework a través d'aquesta classe, sense saber els detalls de funcionament.

Seguidament veurem dos diagrames de seqüència que on es reflexa aquesta arquitectura.

9.4 Inici de partida

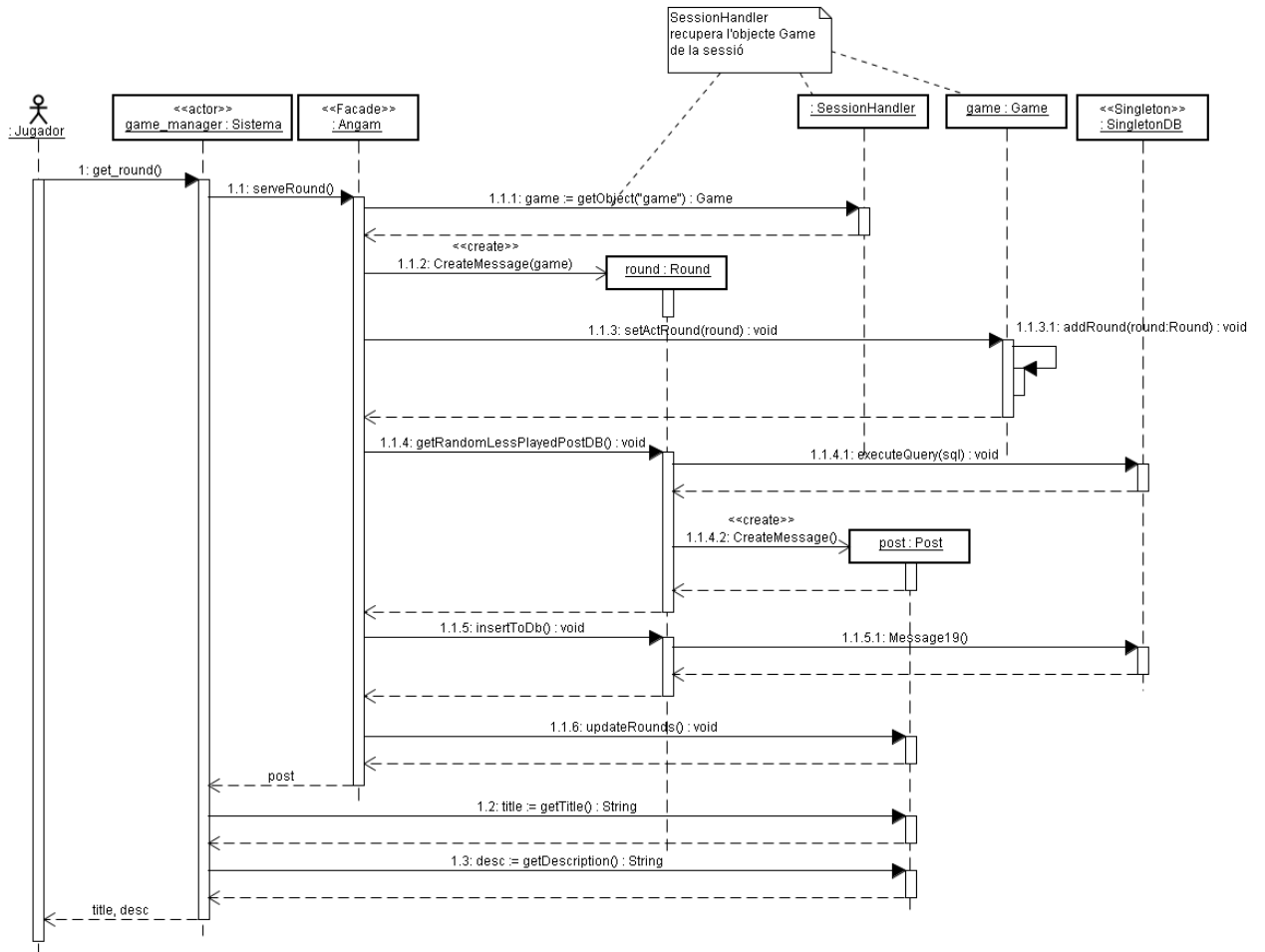
En el següent diagrama podem veure la seqüència d'esdeveniments que es donen quan un jugador inicia una partida.



El jugador fa una petició d'inici de partida sobre el sistema, el qual ve representat per l'script game_manager (codi client que usa el framework), i aquest passa la petició a la classe Angam, la qual implementa la funció que d'inici de partida.

9.5 Servir Round

En el següent diagrama podem veure la seqüència d'esdeveniments que es donen quan un jugador passa a un nou torn de joc i el sistema li ha de servir un nou post.



9.6 Funcionalitats del sistema

Seguidament fem una explicació de les funcionalitats més rellevants que s'han implementat.

9.6.1 Servir post

El sistema s'encarrega de que tots els posts es juguin aproximadament el mateix número de vegades, i a la vegada també s'encarrega de que no sigui sempre amb el mateix ordre. L'algorisme selecciona un post aleatòriament entre els 10 post menys jugats. A la vegada, també té en compte de que no es poden repetir posts. És a dir, no es pot servir un post que ja ha estat servit en la mateixa partida.

Aquesta funcionalitat la implementa el mètode getRandomLessPlayedPostDB() de la classe Round.

9.6.2 Incorporació de posts a la DB

- La incorporació és sota demanda
- Per tal que l'algorisme de servir post funcioni correctament, al incorporar els posts, es té en compte la mitjana de rounds jugats a tots els posts, sinó sempre es servirien els últims post que s'acaben d'incorporar a la base de dades. Aquest algorisme l'implementa la funció insertToDb() de la classe Post
- Al incorporar blogs i posts, primer es comprova si ja existeixen. En cas afirmatiu es fa update

9.6.3 Sistema de puntuació

A l'hora d'assignar punts per les anotacions fetes, els punts es calculen en funció de quants altres jugadors han fet la mateixa anotació. No és un sistema just, ja que la primera vegada que s'anoti un post (correctament), com que la base de dades no tindrà cap referència d'aquell post, donarà zero punts. En canvi el sistema si que és més just quan un post ja s'ha jugat diverses vegades.

Aquesta funcionalitat l'implementa calculatePoints() de la classe Move.

9.7 Persistència

Com ja hem explicat, s'ha decidit fer servir una base de dades relacional per tenir en local les dades dels posts, i a més s'aprofitarà per guardar totes les dades que volem fer persistents.

9.7.1 Definició d'entitats (taules)

```

blog(idblog, dataalta, urlfeed, user_iduser)

game(idgame, idplayer1, ttl1, idplayer2, ttl2)
  {idplayer1} és clau forana de player(idplayer)
  {idplayer2} és clau forana de player(idplayer)

move(idmove, round_idround, value, player_idplayer, points)
  {round_idround} és clau forana de round(idround)
  {player_idplayer} és clau forana de player(idplayer)

player(idplayer, user_iduser)
  {user_iduser} és clau forana de user(iduser)

post(idpost, blog_idblog, permalink, title, description, rounds, passed,disabled)
  {blog_idblog} és clau forana de blog(blog)

round(idround, game_idgame, post_idpost)
  {game_idgame} és clau forana de game(idgame)

user(iduser, username, userpassword)

```

blog

Guarden les dades dels blogs els posts dels quals estem utilitzant a la partida.

post

Guarden les dades dels posts. Els camps `permalink`, `title`, `description` són les dades del post. El camp `rounds`, és un comptador de les vegades que el post s'ha fet servir en alguna partida. El camp `passed` és un comptador de vegades que el post s'ha marcat com 'No té contingut poètic'. El camp `disabled` indica si el post està habilitat per fer-lo servir en les partides. Si un post es marca reiteradament com a 'No té contingut poètic', el sistema acaba desestimant-lo i activant el flag de `disabled`.

move

Guarda les dades de totes les descripcions (anotacions) que es fan del posts. El camp `round_idround` ens diu a quin torn de joc pertany la descripció. El camp `value` ens diu quin valor té la descripció (és la descripció en si). El camp `points` ens diu quants punts ha atorgat el sistema per la descripció.

round

Representa els torns de joc, i ens diu quin post s'ha fet servir i a quina partida.

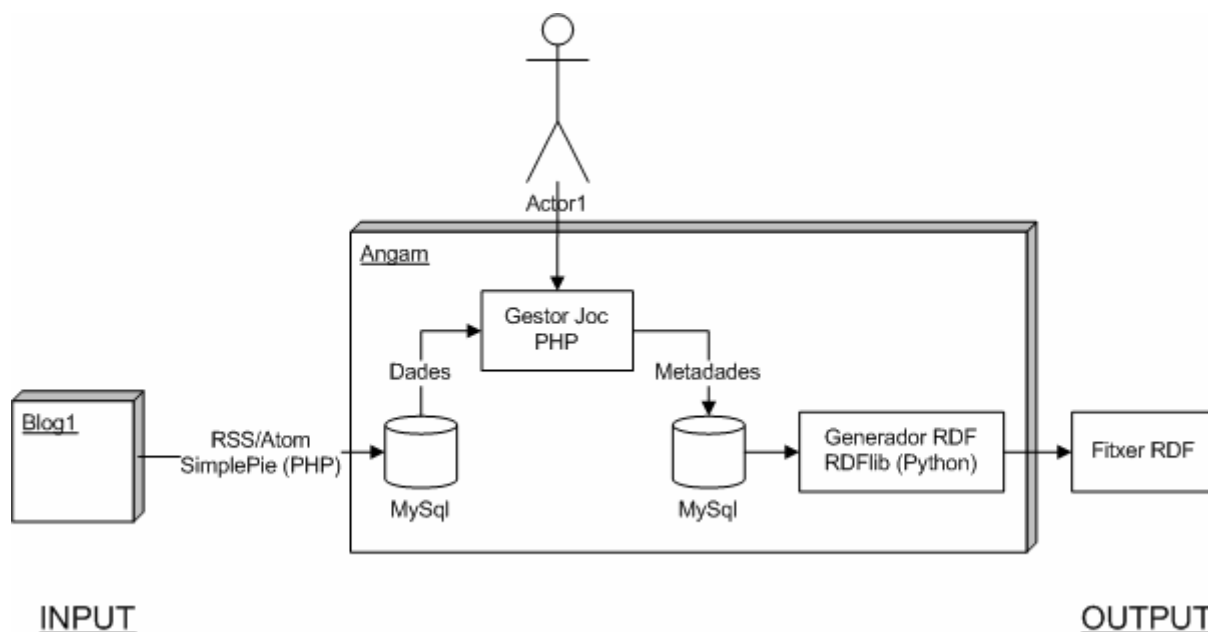
game

Ens permet saber quins jugadors estan assignats a cada partida i els camps t_{t11} (time to live player 1) i t_{t12} (time to live player 2) ens permeten fer el control de si el jugador ha abandonat la partida o no, segons l'algorisme que hem vist anteriorment (vegeu el punt 'Algorisme d'inici de partida').

10 Prototip – Implementació

10.1 Arquitectura

En el següent diagrama podem veure un esquema de l'arquitectura del sistema:



El sistema té una entrada, que són els blogs que volem processar i té una sortida, que és un fitxer de sentències RDF, dels blogs ja processats.

Les dades dels blogs es recullen sota demanda de l'administrador del sistema i es guarden en una base de dades local MySQL. Per fer aquest procés, s'ha fet servir la llibreria de PHP SimplePie .

Un cop les dades estan en el nostre sistema, es fan servir múltiples vegades en diferents partides, i se'n extreuen metadades, les quals també es guarden en la base de dades local.

La implementació de la part del servidor del joc s'ha desenvolupat amb PHP i a la banda client s'ha fet servir la llibreria jQuery per implementar una interfície d'usuari dinàmica, fent ús de tècniques Ajax en la comunicació client - servidor.

Finalment, s'ha implementat un procés per la generació del fitxer RDF, el qual extreu les metadades de la base de dades i les serialitza amb el format RDF/XML. Aquest procés s'ha implementat amb Python fent ús del mòdul RDFlib, i també s'executa sota demanda de l'administrador del sistema.

10.2 Codificació

Tot el codi referent al joc s'ha desenvolupat amb PHP5 sota el paradigma de la orientació a objectes.

Per tal d'accelerar el joc s'ha intentat minimitzar els accessos a la base de dades. Per aconseguir-ho, es fa ús de Sessions de PHP. Concretament es guarden dos objectes a sessió per cada usuari. Un és l'objecte User i l'altre és l'objecte Game al qual està associat el jugador. A la vegada l'objecte Game té referència a tots els objectes Round que s'han jugat, els quals al seu torn tenen referència de tots els objectes Move que s'han fet (anotacions). Això ens permet accedir a totes les dades d'una partida (anotacions, puntuacions, post) sense necessitat d'accedir a la base de dades a cada interacció client – servidor.

L'objecte User és guarda a sessió tant bon punt s'ha loginat correctament, i l'objecte Game just després de crear una partida nova (inici de partida).

La generació del fitxer de sentències RDF s'ha fet amb Python. Inicialment es va provar la llibreria de PHP 'RDF API for PHP' per tal de mantenir una uniformitat en el llenguatge de codificació utilitzat, però les proves fetes amb Python van ser més satisfactòries i finalment es va desestimar la opció amb PHP.

10.3 Components

10.3.1 SimplePie

SimplePie és una llibreria de PHP la qual permet el processat de feeds Atom i RSS indistintament, de manera que no ens cal saber en quin format està el feed.

La llibreria és de lliure distribució i la podem trobar a www.simplepie.org

10.3.2 jQuery

jQuery és una llibreria Javascript, la qual implementa moltes i variades funcions que faciliten la programació d'aplicacions web. En aquest projecte s'ha fet servir bàsicament per que implementa funcionalitats Ajax gràcies a les quals s'ha pogut fer una interfície d'usuari molt dinàmica.

La llibreria és de lliure distribució i la podem trobar a www.jquery.com

10.3.3 RDFlib

RDFlib és una llibreria que permet la manipulació de sentències RDF. Permet tant el "parsejat" com la creació de fitxers RDF amb molta facilitat. A més pot serialitzar amb tots els formats existents avui dia. Nosaltres hem escollit el format RDF/XML.

La llibreria és de lliure distribució i la podem trobar a www.rdflib.net

10.3.4 MySQL

La base de dades relacional que s'ha fet servir és MySQL. Aquesta DB (data base) és, podríem dir, un estàndard d'Internet i té una gran integració amb PHP. A més hi ha disponibles gran quantitat d'eines de lliure distribució que permeten treballar molt còmodament amb la DB.

El programari és de lliure distribució sempre que no sigui per un ús lucratiu i el podem trobar a www.mysql.com

10.4 Fitxer RDF

Les metadades que es generen amb el joc és publiquen mitjançant un fitxer RDF que es genera manualment. S'ha creat un script de Python el qual es connecta amb la base de dades MySQL i selecciona els posts amb els que crearem les sentències RDF.

El següent és un exemple de la sortida de l'script:

```
<?xml version="1.0" encoding="utf-8"?>
<rdf:RDF
  xmlns:sioc='http://rdfs.org/sioc/ns#'
  xmlns:rdf='http://www.w3.org/1999/02/22-rdf-syntax-ns#'
  xmlns:skos='http://www.w3.org/2004/02/skos/core#'
  xmlns:dc='http://purl.org/dc/elements/1.1/'
>
  <sioc:Post rdf:about="http://poesiaula.blogspot.com/2008/11/plom-acer-
reflex-semmiralla-laigua-el.html">
    <dc:title>Va i ve</dc:title>
    <sioc:topic>
      <skos:Concept rdf:about="http://www.game.cat/angam/concepts#mar">
        <skos:prefLabel xml:lang="ca">mar</skos:prefLabel>
      </skos:Concept>
    </sioc:topic>
  </sioc:Post>
  <sioc:Post rdf:about="http://poesiaula.blogspot.com/2008/11/blog-
post_3138.html">
    <dc:title>Lliscar</dc:title>
    <sioc:topic>
      <skos:Concept rdf:about="http://www.game.cat/angam/concepts#equip">
        <skos:prefLabel xml:lang="ca">equip</skos:prefLabel>
      </skos:Concept>
    </sioc:topic>
  </sioc:Post>
</rdf:RDF>
```

11 Conclusions

En base a l'estudi que s'ha fet en el transcurs d'aquest projecte, podem dir que la generació de metadades mitjançant jocs online, és una opció possible per contribuir a solucionar els colls d'ampolla amb els que es troba actualment la web semàntica, per bé que sembla que aquesta no ha de ser la única solució al problema. En aquest sentit, hem vist que també s'està treballant en sistemes d'anotació automàtica.

Actualment hi ha iniciatives de jocs per la generació de coneixement, i si bé sembla que el sector privat hi té interès, la realitat és que la gran majoria són projectes d'investigació científica. Això ens indica que la idea està en una fase molt incipient, que tot just comença.

Per altre banda, no és fàcil crear jocs divertits tant a nivell conceptual com a nivell tècnic, i aquesta és la clau per que el projecte tingui èxit.

De les aplicacions que podem trobar avui dia a la xarxa, la més reeixida sens dubte són els 'Games With A Purpose' de Luis von Ahn, els quals estan servint de model a la majoria de la resta d'aplicacions existents. Qui sap si amb el inici de les primeres anotacions fetes de forma massiva, en breu podríem veure com s'obren altres portes que aclareixin una mica més el futur desenvolupament de la web semàntica.

Hem vist que un dels problemes que té l'anotació de recursos és que calen persones motivades per fer l'anotació. La pregunta que sorgeix ara és: qui crearà aquests jocs i sota quina motivació?