



Ontologia de Twitter: El primer pas per inferir informació dels tweets

Catalina Maria Oliver Vidal
Enginyeria Tècnica en Informàtica de Sistemes

Joan Anton Pérez Braña

07 de gener del 2014



Aquesta obra està subjecta a una llicència de [Reconeixement 3.0 Espanya de Creative Commons](https://creativecommons.org/licenses/by/3.0/es/)

FITXA DEL TREBALL FINAL

Títol del treball:	Ontologia de Twitter: El primer pas per inferir informació dels tweets
Nom de l'autor:	Catalina Maria Oliver Vidal
Nom del consultor:	Joan Anton Perez Braña
Data de lliurament (mm/aaaa):	11/2013
Àrea del Treball Final:	XML i Web semàntica
Titulació:	Enginyeria Tècnica d'Informàtica de Sistemes

Resum del Treball (màxim 250 paraules):

Una ontologia es l'estructura de conceptes d'un domini determinat i la forma en que es relacionen aquests conceptes entre ells.

La representació d'una ontologia ha de mostrar el coneixement o informació que una comunitat de persones té sobre un domini determinat.

La web semàntica ens pot facilitar i agilitzar l'aprenentatge o recerca d'informació a través de les relacions de conceptes que ens aporta gracies a la utilització de les ontologies.

Per la creació de la nostra ontologia hem utilitzat el programa Protégé.

Per el disseny de l'ontologia ens hem basat amb les funcionalitats bàsiques del Twitter pel que hem declarat els conceptes: User, Tweet, Coordinates, Entity, Llista, UserMention, Url, Hashtag, Media, Sizes, Size.

La definició d'aquests conceptes juntament amb les relacions que s'estableixen entre ells ens permeten conèixer el contingut dels tweets, qui els crea, si contenen url's, medias, hashtags, o referències a altres usuaris, saber si aquests usuaris formen part d'una llista o si hi estan subscriptes, saber quins usuaris segueixen o son seguits per altres usuaris, i saber quins tweets son favorits dels usuaris o quins tweets han estat retwittejats per altres usuaris.

Per el poblament de l'ontologia s'ha creat una web amb les llibreries Jena muntat sobre Netbean en Java i amb el servidor Apache Tomcat. Aquesta web ha estat dissenyada amb un front i un back on el front es purament informatiu del proces de poblament i el back es el que realitza tot el tractament del fitxer de l'ontologia i permet introduir-hi les dades desitjades.

Abstract (in English, 250 words or less):

Ontology is an specific domain's concept structure and the way these concepts are related between them.

An ontology must show the knowledge or information that people have about the specific domain.

Semantic web uses ontologies to find information quickly. Semantic Web facilitates learning.

To create our ontology we used Protégé program.

We have used the basic functions of Twitter to design the ontology. Some concepts we use in our ontology are: User, Tweet, Coordinates, Entity, Llista, UserMention, Url, Hashtag, Media, Sizes, Size

Our ontology allows users to know:

- the tweet's contents,
- who created these tweets,
- if the tweets contain URLs, medias, hashtags or references to users,
- if they are part of a list or subscribed to a list,
- which users follow other users and which are followed by others,
- which tweets are the favorites ones and
- which tweets have been retweeted.

A website has been created to insert data into ontology. It uses Jena's libraries and it has been mounted on Netbeans. It works with Java language and Apache Tomcat server.

This website has been designed with a front and a back. The front displays relevant information about the data integration process and the back contains the code for inserting data.

Paraules clau (entre 4 i 8):

Web semàntica, ontologia, Twitter, concepte, relació, propietat.

Índex

1. Introducció.....	2
1.1 Context i justificació del Treball	2
1.2 Objectius del Treball.....	2
1.2.1. Objectius del TFC.....	2
1.2.2. Objectius de l'assignatura.....	3
1.2.3. Coneixements previs.....	3
1.3 Enfocament i mètode seguit.....	3
1.4 Planificació del projecte.....	4
1.4.1. Tasques.....	5
1.4.2. Fites.....	6
1.4.3. Calendari.....	6
1.5 Breu resum de productes obtinguts.....	7
1.6 Breu descripció dels altres capítols de la memòria	7
1.7. Requeriments de maquinari	7
1.8. Requeriments de programari.....	8
1.9. Incidències i riscos – Pla de contingències	8
2. Introducció a l'àrea del TFC: ontologies i web semàntica.....	10
2.1. Ontologies:	10
2.2. Web semàntica:.....	11
2.3. Utilització de les webs semàntiques.....	12
3. Disseny de l'ontologia de Twitter.....	13
3.1. Classe Tweet.....	14
3.2. Classe Coordinates	15
3.3. Classe Users.....	16
3.4. Classe Entities.....	18
3.5. Classe Hashtags	18
3.6. Composició de Media	18
3.8. Classe Sizes.....	20
3.9. Classe URL	20
3.10. Classe UserMention	20
4. Poblament de l'ontologia de Twitter.....	25
4.1. Requeriments	25
4.2. Anàlisi.....	25
4.3. Disseny.....	25
4.4. Implementació	26
5. Conclusions.....	28
6. Glossari	29
7. Bibliografia.....	30
8. Annexos	32
8.1. Annex 1 – Enunciat del treball.....	32
8.2. Annex 2 – Instal·lació del Programari.....	33

1. Introducció

1.1 Context i justificació del Treball

Aquest document recull la memòria del treball de fi de carrera "Ontologia de Twitter: El primer pas per inferir informació dels tweets", per tant hi descriurem els objectius que ens hem proposat assumir, la planificació que considerem oportuna per assumir-ho i la descripció dels passos realitzats juntament amb .

Es tracta d'un treball referent a la investigació de les darreres generacions de web molt oportú amb els temps que corren ja que intentarem donar solució als problemes de comprensió del llenguatge natural humà per par de les computadores, sempre dins el context referent al Twitter i en l'abastiment que ens sigui possible tenint en compte el temps determinat per fer-ho establert per la duració del treball de fi de carrera.

1.2 Objectius del Treball

1.2.1. Objectius del TFC

L'objectiu principal d'aquest treball es crear un sistema que permeti emmagatzemar tweets en ontologies i explotar aquesta informació. Per això ens plantegem una sèrie de subobjectius:

1. Crear o reutilitzar (en cas de trobar-ne alguna amb les característiques que necessitem) una ontologia per emmagatzemar els tweets. Per poder fer això en centrarem amb l'estructura bàsica dels Tweets, la qual podem trobar a l'apartat de desenvolupament de Twitter <https://dev.twitter.com/> i per tant haurem d'aprendre quina es l'estructura en que es basa el Twitter per poder dissenyar la nostra ontologia i aprendre a crear-la.

2. Crear un programa que permeti poblar l'ontologia a partir dels tweets. Una vegada creada l'ontologia l'haurem de poblar d'informació que haurem d'extreure dels tweets i per tant haurem d'investigar com podem poblar l'ontologia d'informació i realitzar les proves pertinents per poder verificar que el procés es realitza de forma correcte i verificar que les dades es guarden de forma correcte.

3. Crear un programa o un conjunt de regles sobre l'ontologia que permeti extreure informació dels tweets de l'ontologia. (sempre que el temps determinat per als dos subobjectius anteriors no s'hagi esgotat). Per poder fer això haurem d'estudiar detalladament la documentació de Twitter i aprendre a utilitzar les seves llibreries per poder crear el programa i aprendre

com connectar-nos al Twitter i així poder extreure la informació que necessitarem per poder utilitzar la nostra ontologia.

1.2.2. Objectius de l'assignatura

Els objectius de l'assignatura son demostrar els coneixements assolits en una seria d'assignatures prèvies demostrant així que al llarg de l'Enginyeria Tècnica en Informàtica de Sistemes hem adquirit els coneixements suficients per poder realitzar un projecte com el que se'ns demana a aquesta assignatura.

Hem de ser capaços:

1. D'analitzar un problema complex i transformar-lo en un projecte informàtic.
2. Planificar i estructurar el desenvolupament del projecte a mitjançant l'elaboració d'un pla de treball.
3. Realitzar recerques d'informació per assolir ràpidament coneixements nous i ser capaços de posar-los en pràctica.
4. Trobar una solució al problema proposat i dur-la a terme d'una forma viable i realista per el temps determinat per realitzar aquesta solució.
5. Elaborar una memòria del projecte amb una estructura prefixada.
6. Elaborar una presentació amb el desenvolupament i els resultats finals del projecte.

1.2.3. Coneixements previs

Tot i que en principi no es necessiten coneixements previs, seria convenient tenir coneixements de xml, d'ontologies i del Twitter en general per poder familiaritzar-se amb facilitat amb la seva API.

Tal i com hem pogut veure en el mòdul 1 del temari també ens hauria estat bastant útil tenir alguns coneixements d'intel·ligència artificial així com de les formes basiques que se solen utilitzar per emmagatzemar coneixements en una base de dades de manera ordenada i optima per les seves consultes.

1.3 Enfocament i mètode seguit

Per realitzar aquest treball ens volíem centrat en l'estudi dels de les ontologies ja realitzades per intentar millorar-les, però ja que no hem trobat

ontologies que puguem millorar per adaptar-les al nostre propòsit ens hem centrat en crear una ontologia que tenint en compte que està feta des de zero i no pot abastar tot el llenguatge natural, establim una limitació del llenguatge utilitzat per intentar assumir la major part del llenguatge que puguem trobar en un tweet.

Per la realització de l'ontologia hem utilitzat el programa protege i utilitzem el format owl. Tenint en compte que no havia utilitzat mai aquest programa i que tampoc havia creat mai cap ontologia s'assumeix que una gran part del temps destinat a la realització de l'ontologia l'he hagut de dedicar a l'aprenentatge tant del programa en si com de la creació de ontologies i a la recerca de informació sobre owl, rdf i rdfschema, sparql i com utilitzar-ho correctament.

Per poder crear un programari que permeti poblar l'ontologia a partir dels tweets he hagut d'estudiar a fons l'api de twitter i revisar els mètodes d'extracció de dades més factible per poder introduir la informació dels tweets a l'ontologia.

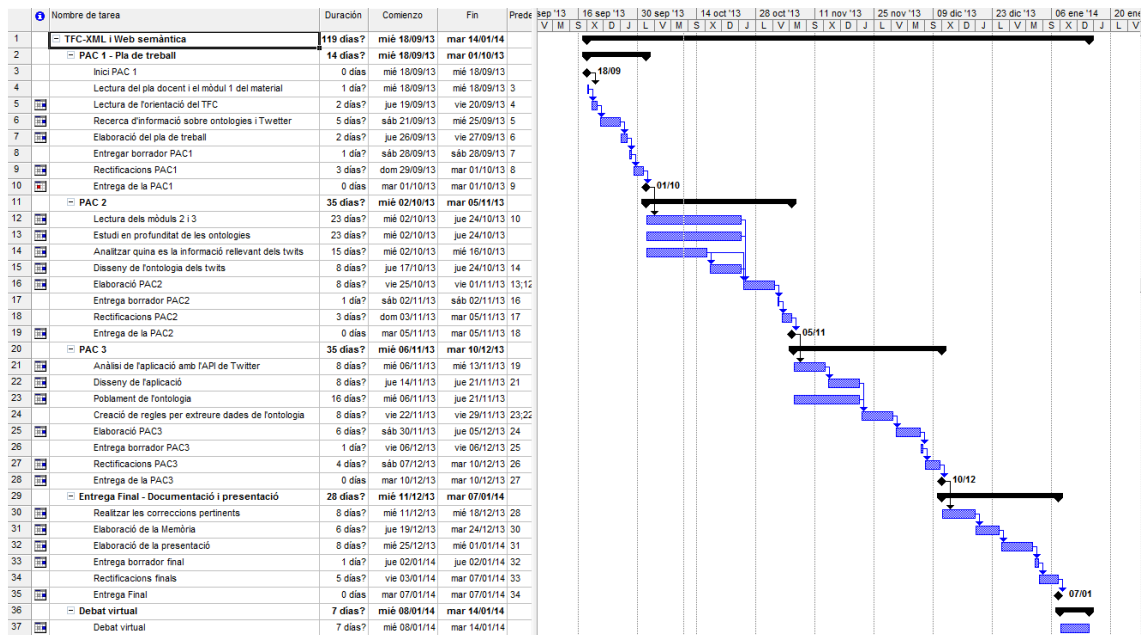
Pel que he pogut veure fins al moment hauré de crear una petita pagina web que hem serveixi de transferència de dades entre el twitter i l'ontologia, aquesta web haurà de contenir la funcionalitat necessària per poder utilitzar-la com a pont de dades.

Hauria de dedicar un cert temps a revisar els programes ja fets d'emmagatzematge de tweets ja que podria intentar reutilitzar-ne algun i d'aquesta manera poder centrar el treball en la creació o millora de l'ontologia ja que considero que es la part més important d'aquest treball i que pel que he pogut comprovar amb la recerca d'informació realitzada aquests dies he pogut comprovar que existeixen varis programes que guarden informació dels tweets com puguin ser Mytwebo, Tweet Backup, ...

1.4 Planificació del projecte

Per la planificació del projecte establirem les fites en les entregues de les PACs establertes per el treball de fi de carrera i per a cadascuna d'aquestes entregues establirem una sèrie de tasques a realitzar.

A continuació indiquem les tasques planificades en el diagrama de Gantt, encara que aquestes tasques només son una aproximació del que s'espera realitzar ja que es possibles que algunes de les tasques variïn en el temps establert i seria possible que s'hagués d'afegir alguna tasca més depenent de la informació trobada en alguna de les tasques d'estudi o recerca:



Per la realització d'aquest diagrama s'ha tingut en compte que només jo faré feina en el treball i per tant he tingut en compte l'horari que tinc disponible. Com es tracte d'una aproximació he tingut en compte que treballaré unes tres hores diàries contant els dies laborals com els caps de setmana, encara que en realitat els dies laborals només podré dedicar una hora o dues com a màxim i la resta d'hores els hauré de recuperar el cap de setmana ja que el meu horari laboral no em permet dedicar més hores al treball de fi de carrera.

1.4.1. Tasques

En aquest apartat indiquem les tasques que tenim previstes:

Tasca del TFC "XML i Web semàntica"	Precedents
1. PAC 1 - Pla de treball	
1.1. Inici PAC 1	
1.2. Lectura del pla docent i el mòdul 1 del material	1.1
1.3. Lectura de l'orientació del TFC	1.2
1.4. Recerca d'informació sobre ontologies i Twitter	1.3
1.5. Elaboració del pla de treball	1.4
1.6. Entregar borrador PAC1	1.5
1.7. Rectificacions PAC1	1.6
1.8. Entrega de la PAC1	1.7
2. PAC 2	
2.1. Lectura dels mòduls 2 i 3	1.6
2.2. Estudi en profunditat de les ontologies	
2.3. Analitzar quina es la informació rellevant dels tweets	
2.4. Elaboració PAC2	2.1;2.2;2.3
2.5. Entrega borrador PAC2	2.4
2.6. Rectificacions PAC2	2.5

2.7. Entrega de la PAC2	2.6
3. PAC 3	
3.1. Anàlisi de l'aplicació amb l'API de Twitter	2.5
3.2. Disseny de l'aplicació	3.1
3.3. Poblament de l'ontologia	
3.4. Creació de regles per extreure dades de l'ontologia	3.2;3.3
3.5. Elaboració PAC3	3.4
3.6. Entrega borrador PAC3	3.5
3.7. Rectificacions PAC3	3.6
3.8. Entrega de la PAC3	3.7
4. Entrega Final - Documentació i presentació	
4.1. Realitzar les correccions pertinents	3.4
4.2. Elaboració de la Memòria	4.1
4.3. Elaboració de la presentació	4.2
4.4. Entrega borrador final	4.3
4.5. Rectificacions finals	4.4
4.6. Entrega Final	4.5
5. Debat virtual	
5.1. Debat virtual	

1.4.2. Fites

Fita	Data
Inici PAC 1	18/09/2013
Entrega de la PAC1	01/10/2013
Entrega de la PAC2	05/11/2013
Entrega de la PAC3	10/12/2013
Entrega Final	07/01/2014

1.4.3. Calendari

Per simplificar el calendari he considerat que treballaré quatre hores diàries tots els dies tant si son dies laborals com els caps de setmana. Evidentment això no es realista ja que els dies laborals no dispo de quatre hores per poder dedicar a la realització del TFC, però tinc previst intentar dedicar el màxim d'hores possibles els dies laborals i els caps de setmana recuperar les hores que hauria d'haver fet els dies laborals.

D'aquesta manera el que faria una setmana seria de dilluns a divendres dedicar unes dues hores diàries i tant el dissabte com el diumenge dedicaria unes nou hores al projecte.

Com ja comentaré en l'apartat d'Incidències i riscos, amb aquestes hores ja tinc en compte que passin els problemes més habituals com problemes de programari o de funcionalitats típiques de tots els projectes en que no es coneix

prèviament el programa ni el llenguatge de programació, i com que en aquest cas no tinc coneixements previs ni del programa, ni del llenguatge ja tinc en compte que passaran i per aquest motiu he posat un nombre d'hores setmanals bastant elevat tenint en compte que treballo de les 9 a les 18 i tinc una hora de desplaçament fins al lloc de treball suposant que no hem facin quedar mes hores per reunions o imprevists a resoldre fora de l'horari laboral.

1.5 Breu sumari de productes obtinguts

Durant el treball de fi de carrera he creat aquesta memòria que conté una explicació del que s'ha anat fent durant tot el procés de realització del treball de fi de carrera juntament amb la recopilació de la informació necessària per poder entendre totes les passes que s'han seguit per la seva realització.

S'han obtingut els fitxers `Twitter.pprj` que es el que correspon al projecte de l'ontologia realitzat amb Protege, i juntament amb aquest el fitxer `Twitter.owl` que es el que correspon a l'ontologia.

Per el poblament de l'ontologia s'ha utilitzat el Netbeans i per tant s'han obtingut tots els fitxers que es creen amb un projecte de Netbeans realitzat en java, dels quals els fitxers principals seran els creats específicament per poblar l'ontologia `index.jsp` amb el contingut web i `Ontologia.java` amb el codi java que dona funcionalitat a la web.

S'ha preparat una presentació en Power Point amb un resum del treball de fi de carrera per una comprensió mes senzilla i visual del projecte.

1.6 Breu descripció dels altres capítols de la memòria

En la resta de capítols fem una petita introducció de les ontologies i les webs semàntiques indicant també quina es la seva utilització.

Indiquem els passos que hem realitzar per dissenyar la ontologia descrivint els conceptes que la formen i com es relacionen.

Descrivim com s'ha realitzat tant la instal·lació com els passos per la programació dels components necessaris per poder poblar la ontologia.

1.7. Requeriments de maquinari

Ordinador amb connexió a Internet per poder realitzar tant la recerca d'informació com les connexions amb Twiter.

1.8. Requeriments de programari

PROTÉGÉ 3.0 (programari gratuït específic per la creació i manipulació d'ontologies amb el qual s'ha creat el projecte Twitter.pprj i la ontologia Twitter.owl amb la que s'ha pogut dissenyar la ontologia bàsica amb les classes i propietats que necessitem per poder guardar el contingut dels Tweets de Twitter i relacionar-los amb els Usuaris i amb els continguts del propi Tweet)

Microsoft Word

Microsoft Project

Microsoft Power Point

1.9. Incidències i riscos – Pla de contingències

Les possibles incidències que penso que poden succeir i el pertinent pla de contingències es descriu a continuació:

1. Desplaçaments per motius de treball

Es poc probable que ocorri aquesta situació ja que generalment els desplaçaments que acostumem a fer no afecten gaire a l'horari de treball perquè quasi tots els clients es troben en la mateixa localitat, però en cas de que m'enviessin de viatge hauria de tenir en compte que em podria trobar en la situació d'estar sense Internet durant uns dies el qual retardaria el treball, en aquest cas no em quedaria més opció que intentar recuperar el temps perdut quan torni tenir Internet i avançar amb el que pugui sense connexió a fi de tenir el menor retard possible.

2. Avaria del punt de treball

En cas de que hi hagués algun tipus d'averia en el punt de treball tinc pensat fer una còpia de seguretat de tots els documents dels que consti el TFC per minimitzar l'impacte i d'aquesta forma no perdre informació o perdre'n el mínim possible. D'aquesta forma evitarem la pèrdua d'informació i en cas de que no es pogués solucionar l'averia en un termini curt sempre es podria intentar avançar d'es d'un altre punt de treball.

3. Problemes de programari

En cas de que hi hagi problemes de programari com puguin ser incompatibilitats de software o com puguin ser problemes de funcions obsoletes tindrem en compte que s'hauran de trobar solucions tant si s'ha de canviar el llenguatge pensat per realitzar l'aplicació com pugui ser la utilització d'alguna aplicació o funcionalitat ja creades per poder enllaçar l'ontologia amb el Twitter. Per tot això ja s'han tingut en compte unes hores a la planificació pensades per poder resoldre problemes de programari al igual que en la cerca d'informació sobre el programari, però en cas de que

no fos suficient s'intentarien dedicar més hores de les establertes per intentar que les dates clau del projecte no sofreixin cap retràs.

4. Problemes de xarxa i/o connexions

En cas de que hi hagués algun problema en la connexió a Internet o amb el Twitter penso que podria afectar en el retràs puntual d'alguna de les tasques previstes i en cas que succeís intentaria aconseguir un altre accés a la xarxa i en cas de que això no fos possible, s'intentarien recuperar les hores del retràs que hagi pogut ocasionar aquest problema.

5. Compromisos

Tinc planejats dos viatges que transcorreren durant l'elaboració del treball, tinc en compte que hem pot suposar un retràs d'uns 5 des en total, però no ho he tingut en compte a la planificació ja que intentaré recuperar les hores que podria perdre d'estudi en aquests viatges abans de realitzar-los de manera que tingui el menor impacte possible en la realització del TFC.

6. Motius de salut

Si per motius de salut es baixa el rendiment , s'intentarà recuperar el màxim nombre d'hores perdudes en la mesura del possible tenint en compte que no puc deixar les hores laborals.

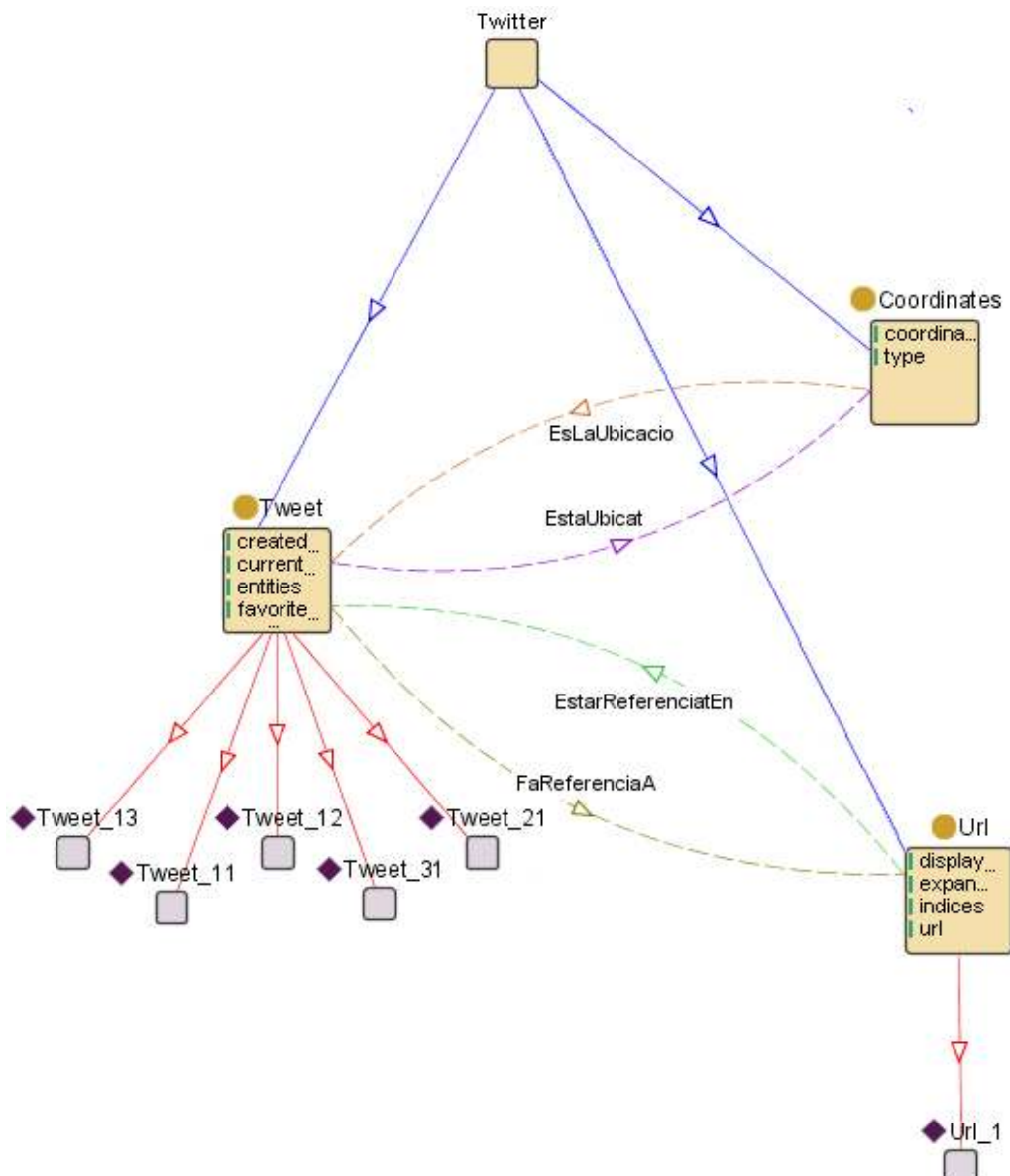
2. Introducció a l'àrea del TFC: ontologies i web semàntica.

2.1. Ontologies:

Les ontologies són la base per poder arribar a la nova generació de webs, les webs semàntiques, que en lloc de funcionar com les webs que coneixem fins a l'actualitat pretenen donar una passa més i fer que les màquines tinguin la capacitat de reconèixer el coneixement a partir de conceptes descrits en les ontologies i d'aquesta forma "entendre" el llenguatge natural.

Les ontologies són imprescindibles per aconseguir la web semàntica ja que la representació d'una ontologia ha de representar el coneixement/informació que una comunitat de persones té sobre un domini, sobre una part de la realitat i que permet la seva interpretació per un sistema informàtic.

Tècnicament, els elements que constitueixen una ontologia són els conceptes, les relacions, les instàncies i els atributs:



- Els conceptes (vèrtex de la xarxa com Tweet, Coordinates o Url de l'esquema anterior) representen les categories ontològiques que són rellevants al domini.
- Les relacions (arcs de la xarxa com EstaUbicat o FaReferenciaA de l'esquema anterior) connecten semànticament els conceptes i indiquen de quina forma es relacionen els conceptes entre ells.
- Les instàncies (vèrtex fulla de la xarxa com Url_1 o Tweet_11 de l'esquema anterior) representen els objectes concrets del domini, vèrtex que es classifiquen per mitjà de les relacions amb els conceptes.
- Els atributs (propietats dels conceptes com url del concepte Url, type del concepte Coordinates o entities del concepte Tweet tal com podem observar en l'esquema anterior) representen les propietats i els seus valors assignats de conceptes o instàncies. Els valors són alhora altres conceptes o instàncies de la mateixa xarxa.

D'aquesta manera una ontologia permet modelitzar relacions molt més complexes.

Per tenir més clar com es defineixen aquestes relacions més complexes hem de tenir en compte que treballem sobre un domini concret i per tant l'abast de l'esquema ontològic queda determinat per els conceptes que haurem determinat que s'inclouen en aquest domini. Per tant podrem tenir major cura amb cadascun dels conceptes i relacions que el formen.

Les relacions son la part fonamental de l'ontologia ja que es la base de la seva funcionalitat, un concepte pot estar relacionat amb altres conceptes o amb altres elements del propi concepte, i pot estar relacionat amb un únic element del concepte relacionat o amb varis elements del concepte relacionat. Això quedarà definit per el rang establert a la relació. Un exemple d'això el podríem tenir amb un concepte "Persona" i una relació "Mare", el concepte Persona estaria relacionat amb ell mateix on el rang indicaria 1 a 1 ja que cada persona només pot tenir una única mare.

Una ontologia defineix els termes a utilitzar per descriure i representar un àrea de coneixement. Les ontologies són utilitzades per les persones, les bases de dades, i les aplicacions que necessiten compartir un domini d'informació

2.2. Web semàntica:

En la web semàntica o web 3.0 veiem que el significat de la informació i el llenguatge natural dels essers humans es fa comprensible per a les màquines.

Les webs semàntiques es defineixen al voltant de entitats i no només paraules, totes les paraules que signifiquen una sola cosa es poden mapejar de manera que aquell significat es presenti com una sola entitat que la maquina es capaç

d'identificar gracies a aquests mapejos i d'aquesta manera les entitats representen coses úniques a la vida real i partir d'aquest moment les maquines comencen a ser capaces de processar el coneixement humà.

La Web semàntica pot ser considerada com una Web de dades, ja que no tant sols aporta estructura visual a la informació sinó que també li proporciona una estructura conceptual que especifica un conjunt de conceptes i defineix les relacions que s'estableixen entre ells (ontologia) – que és capaç de ser compartida, processada i interpretada per agents de programari.

Les webs fins al moment estaven pensades per a les persones, però la web semàntica està pensada per les maquines, perquè les maquines puguin interpretar el seu contingut de manera que puguin identificar si un document es d'un escriptor o si el document parla d'un producte concret etc..

Aquest fet suposa que en una societat on tenim un excés d'informació les maquines suposarien una gran ajuda per poder trobar la informació que busquem de forma molt mes concreta i exhaustiva ja que serien capaces d'interpretar el concepte concret que busquem i mostrar-nos la informació desitjada en el context esperat i per tant les recerques d'informació serien molt mes eficients.

La web semàntica es una web de dades i no de documents com fins ara, no necessita ningú que llegeixi els documents per interpretar-los ja que la pròpia web pot interpretar les dades que conté.

La forma d'entendre les dades no deixa de ser un acord humà, quan el varen crear les web semàntiques es varen haver de posar d'acord en la forma en que es descriurien aquestes dades o conceptes a les maquines perquè tots els conceptes estassin expressats de la mateixa forma, a aquestes formes de descriure les entitats se les va denominar ontologies.

2.3. Utilització de les webs semàntiques

La web semàntica està pensada per fer que els contingut dels recursos de la web, tant siguin imatges com vídeos com textos ..., siguin mes accessibles i mes fàcilment utilitzables pels internautes. La web semàntica integra en la web actual la possibilitat d'agregar varies dades relacionades entre elles.

L'utilització de la web semàntica pot estar relacionada amb molts camps diferents com per exemple en el cap de la investigació ja que pot donar a conèixer molt ràpidament estudis relacionats amb l'àmbit d'interès, publicacions, proves realitzades, etc...

També podria ser de molta utilitat en àmbits d'educació ja que quan es cita un personatge històric el podem relacionar directament amb l'època en la que va viure, les coses que va fer o descobrir, etc...

En general la web semàntica ens pot facilitar i agilitzar l'aprenentatge o recerca d'informació a traves de les relacions de conceptes que ens aporta.

3. Disseny de l'ontologia de Twitter

El disseny d'una ontologia implica fer un estudi per decidir quins conceptes i quines relacions ha de contenir i de quina forma s'han de representar

Determinem el domini que volem cobrir que en el nostre cas son les funcionalitats bàsiques del Twitter, el contingut dels Tweets, qui el crea, si contenen url's, medias, hashtags, o referències a altres usuaris, saber si aquests usuaris formen part d'una llista o si hi estan subscrits, saber quins usuaris segueixen o son seguits per altres usuaris, i saber quins Tweets son favorits dels usuaris.

Una vegada que cercant ontologies ja creades no n'he trobat cap amb el domini que hem establert analitzem les dades que ens ofereix Twitter per determinar els conceptes i les relacions de l'ontologia.

Per realitzar l'ontologia del Twitter ens basarem en l'estructura que ens aporta la documentació de la part de desenvolupaments sobre Twitter. De forma que definirem els conceptes que correspondran a les classes que s'utilitzen en Twitter, al igual que els camps de cada classe i les propietats que indiquen les relacions entre les classes existents en una ontologia, i un cop fet el disseny de les classes i les propietats de l'ontologia procedim a indicar una sèrie d'instàncies de cada classe que indiquen l'individu en si que extraiem de les dades de Twitter per introduir-los en l'ontologia.

Observem les opcions d'usuari per fer-nos una idea de com els utilitza el Twitter i així poder definir millor les classes amb les seves propietats:



Observem que podem veure els Tweets que creem o retwitem com a usuari, els usuaris que ens segueixen, als que seguim, els nostres twits preferits i les llistes a les que ens hem subscrit.

Hem de tenir en compte que retwitejar un tweet no es mes que crear un tweet però indicant que el creador oficial d'aquest tweet es un altre usuari indicant qui es i quan va crear el tweet original del que estem fent una mena de copia.

3.1. Classe Tweet

Els Tweets són l'entitat bàsica dels usuaris de Twitter. Els Tweets poden ser incrustats, resposts, preferits, deixar de ser preferits i poden ser eliminats. Quan diem que poden ser incrustats ens referim a que poden ser creats com a Tweet independent, poden ser la resposta d'un altre Tweet o poden ser retwitejats d'un Tweet ja existent.

Camps de Tweet (propietats)		
Camp	Tipus	Descripció
coordinates	Single Coordinates	Admet valors Null. Representa la ubicació geogràfica d'aquesta Tweet segons l'informat per l'usuari o l'aplicació client . La matriu de coordenades intern té el format GeoJSON (longitud , després latitud) .
created_at	Single String	Hora UTC en què es va crear aquest Tweet.
current_user_retweet	Multiple String	Id del Tweet de l'usuari al qual s'ha retwitejat. Dona suport al paràmetre include_my_retweet, quan s'estableix en true.
entities	Multiple String	Entitats que han estat analitzades pel text del tweet . Les entitats del Tweet poden ser urls, hashtags i media files com fotografies.
favorite_count	Single Int	Admet valors Null. Indica aproximadament quantes vegades aquest Tweet ha estat indicat com a "preferit" pels usuaris de Twitter.
favorited	Single Boolean	Admet valors Null. Indica si aquest Tweet ha estat indicat com a preferit per l'usuari que s'autentica.
filter_level	Single String	Indica el valor màxim del paràmetre de filter_level que pot ser utilitzada per aquest Tweet. Els valors que pot tenir son: "none", "low", o "medium", es reserva "high" per possibles usos futurs. Així, el valor de "medium" serà com la transmissió d'un "none".
id	Single Int	La representació sencera de l'identificador únic per aquest Tweet. Aquest nombre és major que 53 bits i alguns llenguatges de programació poden tenir dificultats en la seva interpretació. L'ús d'un sencer amb signe de 64 bits per emmagatzemar aquest identificador és segur. S'ha d'utilitzar "id_str".
id_str	Single String	La representació de cadena de l' identificador únic per aquest Tweet. Les implementacions han d'utilitzar aquest en lloc del "id".
in_reply_to_screen_name	Single String	Admet valors Null. Si el Tweet representat és una resposta, aquest camp conté el nom d'usuari de l'autor del tweet original.

in_reply_to_status_id	Single Int	Admet valors Null. Si el Tweet representat és una resposta, aquest camp conté el nombre sencer de l'identificació del Tweet original.
in_reply_to_status_id_str	Single String	Admet valors Null. Si el Tweet representat és una resposta, aquest camp conté la cadena de l'ID del Tweet original.
in_reply_to_user_id	Single Int	Admet valors Null. Si el Tweet representat és una resposta, aquest camp conté el nombre sencer de l'ID de l'autor original del tweet. Això no vol dir necessàriament que sempre sigui l'usuari esmentat directament al Tweet.
in_reply_to_user_id_str	Single String	Admet valors Null. Si el Tweet representat és una resposta, aquest camp conté la cadena que representa l'ID de l'usuari que es l'autor original del Tweet. Això no vol dir necessàriament que sempre sigui l'usuari esmentat directament al Tweet.
lang	Single String	Admet valors Null. Quan aquest camp té valor diferent de Null, indica l'identificador de l'idioma del Tweet.
retweet_count	Single Int	Nombre de vegades que s'ha retweetejat el Tweet. Aquest camp no té un límit de 99 i no es convertirà en una cadena "100+"
retweeted	Single Boolean	Indica si aquest tweet ha estat retweetejat per l'usuari que s'autentica.
retweeted_status	Single String	Els usuaris poden ampliar la difusió dels tweets publicats per altres usuaris amb el retweeteig. Els retweets es distingeixen dels tweets típics per l'existència de l'atribut retweeted_status. Aquest atribut conté una representació del tweet original que va ser retweetejat. S'ha de tenir en compte que els retweets de retweets no mostren representacions de l'intermediari del retweet, només el tweet inicial. (Els usuaris també poden desretweetejar un retweet que van crear mitjançant la supressió del seu retweet.)
source	Single String	Utilitat utilitzada per publicar el tweet, com una cadena en format HTML. Els tweets des del lloc web de Twitter tenen un valor de codi per la web.
text	Single String	L'actualització d'estat en text format UTF-8.
truncated	Single Boolean	Indica si el valor del paràmetre text es trunca, per exemple, com a resultat d'un Retweet inferior o igual a la longitud de 140 caràcters del Tweet. Un text truncat acabarà en punts suspensius , així ... La gran majoria dels Tweets tindrà aquest camp a false. S'ha de tenir en compte que mentre retweets nadius poden tenir la seva propietat text escurçat, el text original estarà disponible en l'objecte retweeted_status i el paràmetre truncated s'estableix en el valor de la condició original (en la majoria dels casos, false).
user	Single String	L'usuari que va publicar aquest Tweet.

3.2. Classe Coordinates

Coordinates indica les coordenades de la localització en la que es troba l'usuari que escriu el Tweet en el moment de la creació d'aquest.

Camps de Coordenates		
Camp	Tipus	Descripció
coordinates	Multiple <u>String</u>	La longitud i la latitud de la ubicació del tweet, com una col·lecció en forma de [longitud, latitud].
type	Single String	El tipus de dades codificades en la propietat coordinates.

3.3. Casse Users

Els usuaris poden ser qualsevol persona o qualsevol entitat. El usuaris son els que utilitzen les funcionalitats de Twitter, segueixen o son seguits per altres usuaris, creen llistes o es subscriuen a llistes, poden escriure o retwittejar tweets i indicar tweets com a favorits.

Camps de User		
Camp	Tipus	Descripció
created_at	Single String	La data i hora UTC en què el compte d'usuari es va crear el Twitter.
default_profile	Single Boolean	Quan és true, indica que l'usuari no ha alterat el tema o el fons del seu perfil d'usuari.
default_profile_image	Single Boolean	Quan és true, indica que l'usuari no ha pujat el seu propi avatar i es en el seu lloc s'utilitza per defecte un avatar en forma d'ou.
description	Single String	Admet valors Null. Definit per l'usuari en format UTF-8 string descrivint el seu compte.
entities	Multiple String	Entitats que han estat analitzats pels camps d'URL o descripció definida per l'usuari.
favourites_count	Single Int	El nombre de tweets d'aquest usuari ha marcat com a preferits en la vida del compte.
follow_request_sent	Single Boolean	Admet valors Null. Quan és true, indica que l'usuari d'autenticació ha emès una sol·licitud per seguir aquest compte d'usuari protegit.
followers_count	Single Int	El nombre de seguidors que aquest compte té actualment. Sota certes condicions de pressió, aquest camp indica temporalment "0".
friends_count	Single Int	El nombre d'usuaris que aquest compte està seguint. Sota certes condicions de pressió, aquest camp indica temporalment "0".
geo_enabled	Single Boolean	Quan és true, indica que l'usuari ha permès la possibilitat de geolocalitzar els seus Tweets. Aquest camp ha de ser cert per a l'usuari actual per adjuntar les dades geogràfiques utilitzant estats de post / actualització.
id	Single Int	El nombre sencer de l'identificador únic per aquest usuari. Aquest nombre és major que 53 bits i alguns llenguatges de programació poden tenir dificultats en la seva interpretació. L'ús d'un sencer amb signe de 64 bits per emmagatzemar aquest identificador és segur. S'ha d'utilitza id_str per extreure l'identificador.
id_str	Single String	La cadena de l' identificador únic per aquest usuari. Les implementacions han d'usar aquest

		camp en lloc de "id".
lang	Single String	El codi BCP 47 per l'idioma auto-declarat de la interfície de l'usuari. Pot o no tenir res a veure amb el contingut dels seus tweets.
listed_count	Single Int	El nombre de relacions públiques de les que aquest usuari és membre.
location	Single String	La ubicació definida per l'usuari per al perfil d'aquest compte. No és necessàriament un lloc analitzable. Aquest camp de tant en tant es interpretat vagament pel servei de cerca.
name	Single String	El nom de l'usuari, tal com l'han definit. No necessàriament ha de ser el nom d'una persona. En general té un límit de 20 caràcters, però pot estar subjecte a canvis.
profile_background_color	Single String	El color hexadecimal triat per l'usuari per el seu fons de pantalla.
profile_background_image_url	Single String	Adreça URL basada en HTTP que apunta a la imatge de fons carregada per el seu perfil.
profile_background_image_url_https	Single String	Adreça URL basada en HTTPS que apunta a la imatge de fons que l'usuari ha pujat al seu perfil.
profile_background_tile	Single Boolean	Quan és true, indica que la imatge indicada a profile_background_image_url s'ha de mostrar en mosaic quan es carrega ja qe així ho ha indicat l'usuari.
profile_banner_url	Single String	La direcció URL basada en HTTPS que apunta a la representació de la web estàndard del banner del perfil de l'usuari. Mitjançant l'addició d'un element al final de la ruta de l'adreça URL, pot obtenir diferents mides optimitzades per a pantalles específiques.
profile_image_url	Single String	Una adreça URL basada en HTTP que apunta a la imatge del avatar de l'usuari.
profile_image_url_https	Single String	Una adreça URL basada en HTTPS assenyalant l'imatge d'avatar de l'usuari.
profile_link_color	Single String	El color hexadecimal que l'usuari ha optat per mostrar els vincles en la seva interfície d'usuari de Twitter.
profile_sidebar_border_color	Single String	El color hexadecimal que l'usuari ha optat per mostrar les vores laterals amb en la seva interfície d'usuari de Twitter.
profile_sidebar_fill_color	Single String	El color hexadecimal que l'usuari ha triat per mostrar a la barra lateral amb els fons en la seva interfície d'usuari de Twitter.
profile_text_color	Single String	El color hexadecimal que l'usuari ha triat per mostrar el text en la seva interfície d'usuari de Twitter.
profile_use_background_image	Single Boolean	Quan és true, indica que l'usuari vol que la seva imatge de fons pujada sigui utilitzada.
protected	Single Boolean	Quan és true, indica que l'usuari ha optat per protegir els seus tweets.
screen_name	Single String	El nom de pantalla, o àlies amb el que aquest usuari s'identifica. screen_names és únic, però subjecte a canvis. S'utilitza id_str com un identificador d' usuari sempre que sigui possible. Normalment, aquest camp té un màxim de 15 caràcters de longitud, però alguns relats històrics poden existir amb longituds més llargues.
status	Single String	Admet valors Null. Si és possible, el tweet o retweet més recent de l'usuari. En alguns casos,

		aquestes dades no es poden proporcionar i aquest camp tindrà valor null o buit.
statuses_count	Single Int	El nombre de tweets (incloent retweets) expedits per l'usuari.
time_zone	Single String	Admet valors Null. Una cadena que descriu la zona horària que l'usuari declara.
url	Single String	Admet valors Null. Un URL proporcionada per l'usuari associat al seu perfil.
utc_offset	Single Int	Admet valors Null. El desplaçament de GMT / UTC en qüestió de segons.
verified	Single Boolean	Quan és true, indica que l'usuari té un compte verificat.
withheld_in_countries	Single String	Quan te valor, indica amb codis de país de dues lletres on es troba l'usuari.

3.4. Classe Entities

Les Entities proporcionen metadades i informació contextual addicional sobre el contingut publicat a Twitter. Les entities no estan separades dels continguts que descriuen. Les entities juguen un paper decisiu en la resolució de les adreces URL.

3.5. Classe Hashtags

El hashtag es un nom o frase que comença amb el caràcter # i per una llei no escrita si inclou més d'una paraula s'escriurà tot seguit sense separar les paraules amb espais en blanc, el primer caràcter de cada paraula s'escriurà en majúscules.

Camps de Hashtags		
Camp	Tipus	Descripció
indices	Multiple Int	Una matriu d'enters que indiquen els desplaçaments dins del text del Tweet on el hashtag comença i acaba. El primer nombre enter representa la ubicació del caràcter # a la cadena del text del Tweet. El segon nombre enter representa la ubicació del primer caràcter després del hashtag. Per tant, la diferència entre els dos nombres serà la longitud del nom del hashtag més un (pel caràcter '#').
text	Single String	Nom del hashtag, menys el caràcter principal "#".

3.6. Composició de Media

El media es un fitxer que es pot adjuntar al tweet, generalment es tracta d'una fotografia però també es pot tractar d'un vídeo, document, etc ...

Camps de Media		
Camp	Tipus	Descripció
display_url	Single String	URL dels mitjans de comunicació per mostrats als clients.
expanded_url	Single String	Una versió ampliada de display_url. Els enllaços a la pàgina de visualització dels mitjans de comunicació.
id	Single Int	Identificació dels mitjans de comunicació, expressats com un enter de 64 bits.
id_str	Single String	Identificació dels mitjans expressats en forma de cadena.
indices	Multiple Int	Una matriu d'enters que indiquen els desplaçaments dins del text del tweet on l'URL comença i acaba. El primer sencer representa la posició del primer caràcter de la URL al text del tweet. El segon nombre enter representa la ubicació del primer caràcter que no és de l'enllaç de la URL (o al final de la cadena si la URL és l'última part del text del tweet).
media_url	Single String	Una URL http:// que apunta directament a l'arxiu multimèdia carregat.
media_url_https	Single String	Una URL https:// apuntant directament a l'arxiu multimèdia carregat per pàgines https.
sizes	Multiple Sizes	Un objecte que mostra les mides de l'arxiu multimèdia.
source_status_id	Single Int	Per tweets que contenen els mitjans de comunicació diferents als que es van associar inicialment amb un tweet, aquest ID assenyala l'original Tweet.
source_status_id_str	Single String	Per tweets que contenen els mitjans de comunicació diferents als que es van associar inicialment amb un tweet, aquesta cadena a base de punts d'identificació per a l'original Tweet.
type	Single String	Tipus dels fitxers enviats.
url	Single String	URL per l'enllaç dels mitjans de comunicació. Això es correspon amb la URL incrustada directament en el text del tweet, i els valors per als paràmetres d'índexs.

3.7. Classe Size

El size es na propietat que indica el tamany en píxels en el que es desitja que aparegui un media.

Camps de Size		
Camp	Tipus	Descripció
h	Single Int	Alçada en píxels d'aquesta mida.
resize	Multiple String	Mètode de canvi de mida que permet obtenir aquesta mida. Un valor d'ajust que significa que el mitjà es canvia de mida

		per adaptar-se a una dimensió, mantenint la seva relació d'aspecte original. Un valor de la collita significa que els mitjans de comunicació es retallen per tal d'adaptar-se a una resolució específica.
w	Single Int	Amplada en píxels d'aquesta mida.

3.8. Classe Sizes

El sizes indica els possibles dimensionaments establerts per un media, per cadascun d'aquests li correspondrà un size que indicarà la seva mida concreta en píxels.

Camps de Sizes		
Camp	Tipus	Descripció
thumb	Multiple String	Informació per a una versió en miniatura de la mida dels mitjans de comunicació.
large	Multiple String	Informació per a una versió de grans dimensions dels mitjans de comunicació.
medium	Multiple String	Informació per a una versió de mida mitjana dels mitjans de comunicació.
small	Multiple String	Informació per a una versió de mida petita dels mitjans de comunicació.

3.9. Classe URL

La url es una adreça electrònica.

Camps de URL		
Camp	Tipus	Descripció
display_url	Single String	Versió de la URL per mostrar als clients.
expanded_url	Single String	Versió ampliada de display_url.
indices	Multiple String	Una matriu d'enters que representen els desplaçaments dins del text del tweet on l'URL comença i acaba. El primer sencer representa la posició del primer caràcter de la URL al text del tweet. El segon nombre enter representa la ubicació del primer caràcter que no correspon a l'enllaç de la URL ja que s'ha arribat al seu final.
url	Single String	URL corresponent al valor incorporat directament al text del tweet i els valors dels paràmetres indices.

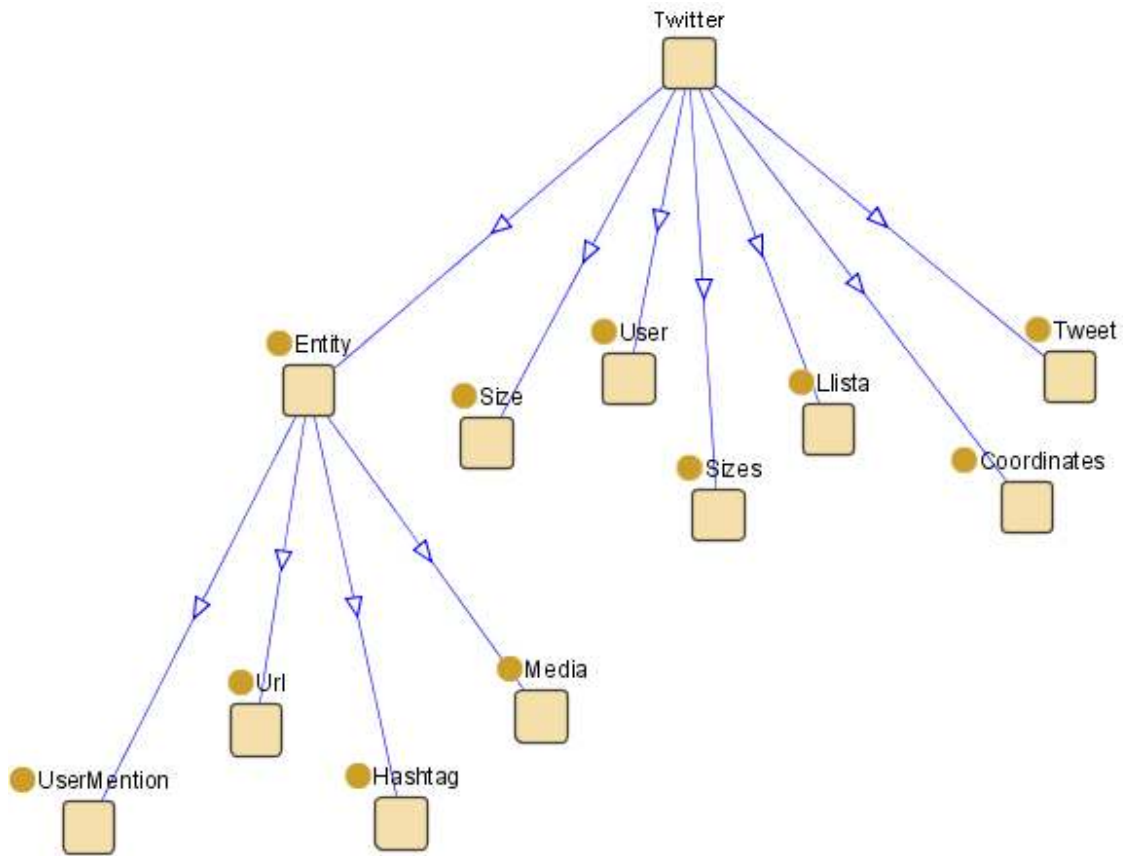
3.10. Classe UserMention

El user mention es la referència que es fa a un usuari concret indicant el seu nom que començarà amb el caràcter @.

Camps de UserMention		
Camp	Tipus	Descripció
id	Single Int	ID de l'usuari que s'ha esmentat, com un enter.
id_str	Single String	ID de l'usuari que s'ha esmentat, com una cadena.
indices	Multiple String	Una matriu d'enters que representen els desplaçaments dins del text del tweet on comença i on acaba la referència per a l'usuari. El primer nombre enter representa la ubicació del caràcter "@" de la menció de l'usuari. El segon nombre enter representa la posició del primer caràcter ja no representa la menció de l'usuari.
name	Single String	Mostra el nom de l'usuari al que es fa referència.
screen_name	Single String	Nom en pantalla de l'usuari al que es fa referència.

A partir d'aquestes dades i tenint en compte les opcions que observem des del punt de vista de l'usuari ens fem un esquema de com relacionar-les per poder construir l'ontologia.

L'esquema basic de classes que ens fem per poder construir la ontologia seria el següent:



D'aquesta forma tenim el concepte de Llista (List) que reuneix un conjunt d'usuaris que pretenen redactar una sèrie de Tweets d'un tema comú i per tant els usuaris poden formar part de la llista o simplement subscriure's per rebre la informació del que s'informa als tweets que s'escriuen per aquesta llista.

També tenim el concepte d'Usuari (User) que es la persona o entitat que realitza les accions dins Twitter, redacta o retwiteja els tweets, es pot fer seguidor d'altres usuaris, pot indicar quins son els seus tweets preferits i pot fer referència a altres usuaris, url's, a un tema concret indicat per un hashtag o inserir medias com imatges al tweets que redacti. El nom dels usuaris comença amb el símbol @ per identificar que es tracte d'un usuari.

El concepte UserMention que es la referència a un usuari feta en un tweet.

El concepte de Hashtag es un tema o tendència referenciada per un nom que comença amb el símbol # i no te espais en blanc, si es volen separar paraules s'escriuen en minúscules a excepció de la primera lletra de cada paraula que comença en majúscula, encara que sigui una regla no escrita.

El concepte Media pot ser tant una imatge, com un vídeo, com una gravació de so, però no distingirem en el tipus del que es tracta ja que en les dades de Twitter ho guarda com al mateix concepte i els distingeix per el camp type.

El concepte sizes que son les dimensions associades a un media per poder-ho mostrar a la web.

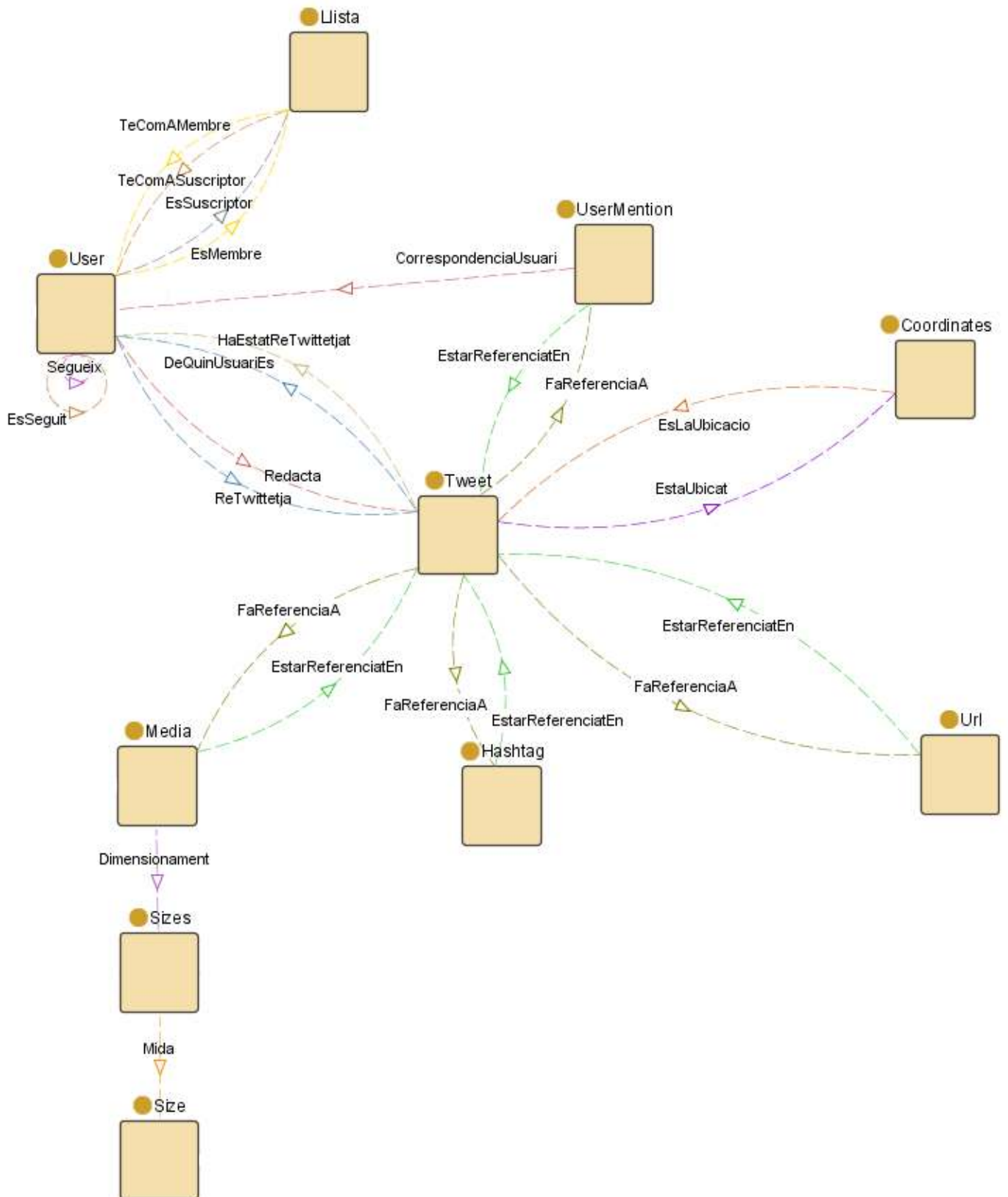
El concepte size es la mida de cada dimensió associada a un media.

El concepte Url conté la adreça web a la que s'hagi fet referència a un tweet.

El concepte Tweet correspon a l'escrit que hagi creat un usuari i que pot fer referència a altres usuaris, hashtags, url's i pot contenir medias.

El concepte Coordinates indica la localització que es pot indicar a un tweet per poder saber des de on s'escriu es tweet.

En el següent diagrama observem els conceptes del disseny de la ontologia juntament amb les relacions semàntiques que els uneixen:



4. Poblament de l'ontologia de Twitter

4.1. Requeriments

Per poder poblar l'ontologia amb les dades de Twitter necessitarem un programa que ens permeti inserir dades de forma massiva a l'ontologia. Per tant tenim els requeriments d'un programa que ens permeti manipular l'ontologia, obrir el fitxer owl, inserir les dades que nosaltres desitgem com a instàncies dels conceptes definits i en les propietats d'aquests conceptes.

4.2. Anàlisi

Per la necessitat de manipular el fitxer de l'ontologia veig necessària la utilització de les llibreries JENA ja que tenen una sèrie de funcions específiques per el tractament d'ontologies i per tant ens facilitarà molt el tractament del fitxer, les insercions de dades, etc..

Ja que el programa a desenvolupar es tracta d'una web per facilitar l'interconnexió amb Twitter hem de tenir en compte que caldrà un entorn apropiat per la creació d'una web per tant utilitzarem l'Apache Tomcat com a servidor web i per la seva compatibilitat amb Netbeans utilitzarem aquest com entorn de programació així que el llenguatge de programació serà Java per ser un dels més còmodes d'utilitzar sobre Netbeans.

Per la manipulació del codi tindrem una part Front on tindrem el contingut web que es mostrarà i una part Back on es realitzarà tota la manipulació de l'ontologia.

4.3. Disseny

En la part Front tindrem una web molt bàsica que simplement tindrà títol i molt poc de disseny ja que es tracta d'una web simplement informativa de la funcionalitat de la part Back per saber si el procés es realitza de forma correcte o es produeix algun error durant el procediment de poblar l'ontologia. Tot això es realitza al fitxer index.jsp

La part Back es l'encarregada de dur a terme tota la manipulació de l'ontologia on per una part s'obre el fitxer owl s'en fa una lectura indicant les classes o conceptes que s'hi troben definits i indicant les instàncies que conté el propi fitxer, per una altra part s'hi troben les dades que s'han d'introduir en l'ontologia i el procés inserció d'aquestes amb les propietats oportunes.

4.4. Implementació

Per realitzar el poblament de l'ontologia hem muntat un entorn apropiat per poder crear una pàgina web capaç d'obrir el fitxer de l'ontologia i poblar-la.

Per muntar l'entorn hem utilitzat una maquina virtual amb el VirtualBox. A aquesta maquina hi hem posat un Windows XP com a sistema operatiu, hi he realitzat la instal·lació de Apache Tomcat per poder utilitzar-lo de servidor web i per l'entorn de programació hem realitzat la instal·lació de Java juntament amb Netbeans que ens permet indicar que volem utilitzar els servidor Apache Tomcat durant la instal·lació.

Al tenir l'entorn muntat procedim a començar amb la programació. Per poder manipular amb major facilitat l'ontologia utilitzem les llibreries.

Per instanciar la ontologia utilitzem les següents instruccions:

```
OntModel ontmodel = null;  
ontmodel = ModelFactory.createOntologyModel(OntModelSpec.OWL_MEM);
```

Per obrir el fitxer amb l'ontologia:

```
InputStream inst =  
FileManager.get().open("http://localhost:8084/poblamentOntologia/Twitter.owl");  
if (inst == null){  
    throw new IllegalArgumentException("Fitxer no trobat");  
}
```

Llegim el fitxer amb la instrucció:

```
ontmodel.read(inst, "");
```

Declarem les classes de la nostra ontologia de la forma següent, com totes les classes es declaren de la mateixa forma indicarem la declaració d'una classe fictícia enomanada nomClasse:

```
OntClass nomClasse = ontmodel.getOntClass("http://www.owl-  
ontologies.com/Ontology1380644855.owl#nomClasse");
```

Declarem el valor individual que volem introduir en aquesta classe:


```
Individual valIndividual = ontmodel.createIndividual("http://www.owl-  
ontologies.com/Ontology1380644855.owl#valIndividualNomClasse");
```

Declarem les propietats de les classes de la forma següent, a l'exemple indiquem com a nom de la propietat prop:

```
OntProperty prop = ontmodel.getOntProperty("http://www.owl-  
ontologies.com/Ontology1380644855.owl#prop");
```

Poblem la instància de la propietat amb la instrucció següent:

```
valIndividual.setPropertyValue(prop,  
ontmodel.createTypedLiteral("valorPropietat"));
```



Poblament de l'ontologia de Twitter

Classes: User, Sizes, Coordinates, Llista, Url, Media, Entity, Size, UserMention, Hashtag, Tweet,

Instàncies: User_3, User_1, User_2, Llista_1, Url_1, Media_1, Hashtag_1, Tweet_31, Tweet_11, Tweet_21, Tweet_13, Tweet_12,

S'ha carregat la instància CoordinatesProva1 de la classe Coordinates

S'ha carregat la instància HashtagProva1 de la classe Hashtag

S'ha carregat la instància MediaProva1 de la classe Media

S'ha carregat la instància UrlProva1 de la classe Url

S'ha carregat la instància UserMentionProva1 de la classe UserMention

S'ha carregat la instància LlistaProva1 de la classe Llista

S'ha carregat la instància SizeProva1 de la classe Size

S'ha carregat la instància SizesProva1 de la classe Sizes

S'ha carregat la instància TweetProva1 de la classe Tweet

S'ha carregat la instància UserProva1 de la classe User

Classes: User, Sizes, Coordinates, Llista, Url, Media, Entity, Size, UserMention, Hashtag, Tweet,

Instàncies: UserProva1, User_3, User_1, User_2, SizesProva1, CoordinatesProva1, LlistaProva1, Llista_1, UrlProva1, Url_1, MediaProva1, Media_1, SizeProva1, UserMentionProva1, HashtagProva1, Hashtag_1, Tweet_31, Tweet_11, Tweet_21, Tweet_13, TweetProva1, Tweet_12,

5. Conclusions

Penso que es un treball que pot esser de molta utilitat per l'avanç de les webs de darrera generació encara que per el temps disponible de la realització del treball només pot ser una petita aportació per avançar en que les maquines siguin capaces d'interpretar el llenguatge natural dels humans.

Amb aquest treball les maquines ja seran capaces d'interpretar les dades que s'utilitzen en Twitter.

6. Glossari

TFC: Teball de Fi de Carrera.

7. Bibliografia

La bibliografia recomanada per la realització del treball es la següent:

- Els moduls del temari:

http://materials.cv.uoc.edu/continguts/PID_00199493/index.html?ajax=true

- Com a guia per realitzar el treball de fi de carrera:

<http://slidesha.re/fty0Rk>

- Sobre W3C

<http://www.w3.org/Home/Lassila>.

- Sobre ontologies i owl:

<http://www.w3.org/2007/09/OWL-Overview-es.htm>

<http://www.slideshare.net/fulvio.corno/ontologies-introduction-designlanguages-and-tools>

- SlideShare recomanats :

<http://www.slideshare.net/snyderp/intro-to-the-semantic-web-peter-snyder-csg339-northeastern-university-1697429>

<http://www.slideshare.net/LeeFeigenbaum/intro-to-the-semantic-web-landscape-2011>

- Protégé:

<http://protege.stanford.edu/>

<http://protege.stanford.edu/download/registered.html>

- Plataforma Twitter:

Tweets: <https://dev.twitter.com/docs/platform-objects/tweets>

Users: <https://dev.twitter.com/docs/platform-objects/users>

Entities: <https://dev.twitter.com/docs/platform-objects/entities>

Places: <https://dev.twitter.com/docs/platform-objects/places>

- Projecte de referència

<http://openaccess.uoc.edu/webapps/o2/handle/10609/18961>

- Elements per implementar l'aplicació:

Twitter Cards: <https://dev.twitter.com/docs/cards>
Embedded Timelines: <https://dev.twitter.com/docs/embedded-timelines>
Twitter for Websites: <https://dev.twitter.com/docs/twitter-for-websites>
Integrating with Twitter on iOS: <https://dev.twitter.com/docs/ios> (no recomanable)

- API de Twitter

Using the Twitter Search API: <https://dev.twitter.com/docs/using-search>
REST API v1.1 Resources: <https://dev.twitter.com/docs/api/1.1>
The Streaming APIs: <https://dev.twitter.com/docs/streaming-apis>
Exploring the Twitter API: <https://dev.twitter.com/console>
Twitter Libraries: <https://dev.twitter.com/docs/twitter-libraries>

8. Annexos

8.1. Annex 1 – Enunciat del treball

Títol

ONTOLOGIA DE TWITTER: EL PRIMER PAS PER INFERIR INFORMACIÓ DELS TWEETS

Descripció del treball

Fer servir ontologies per emmagatzemar la informació rellevant dels tweets dels usuaris i extreure'n informació.

Antecedents

Avui en dia les xarxes socials estan tenint un gran impacte, no només en la manera en com ens relacionem, sinó en tots els aspectes possibles. Les opinions dels usuaris es tenen en compte ara per les empreses alhora de veure quines potencials millores poden, pels governs per copsar la intenció de vot e interessos de la població, pels sociòlegs perquè poden donar pistes sobre els trets característics de la mateixa, etc. Per tant, hi ha un gran interès en poder emmagatzemar i analitzar la informació que viatja al llarg de la xarxa.

Twitter no és una excepció, i té algunes característiques que el fan especialment interessant: missatges curts amb molt contingut i poca opció a utilitzar recursos lingüístics difícils d'interpretar, referències a d'altres tweets i URLs, etc. De fet en els últims anys, hi ha hagut molts esforços en analitzar i extreure informació dels tweets dels usuaris.

Objectius del treball

L'objectiu d'aquest treball es crear un sistema que permeti emmagatzemar tweets en ontologies i explotar aquesta informació. Per tant, el treball tindrà els següents objectius:

1. Crear una ontologia per emmagatzemar els tweets (també es podrà reutilitzar ontologies existents si es troba adient)
2. Crear un programa que permeti poblar l'ontologia a partir dels tweets.
3. Crear un programa o un conjunt de regles sobre l'ontologia que permeti extreure informació dels tweets de l'ontologia.

Els subobjectius 1 i 2 seran d'obligat compliment. El tercer objectiu serà opcional.

8.2. Annex 2 – Instal·lació del Programari

Instal·lació de Protege:

Descarreguem el programari Protégé Desktop 3.5 de la web (<http://protege.stanford.edu/download/registered.html>) tal i com es recomenava en les orientacions dels TFC:

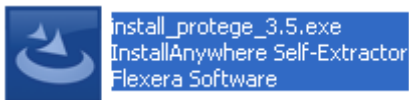


The screenshot shows a web browser window with the address bar containing `protege.stanford.edu/download/registered.html`. The page features the Protege logo and a navigation menu with links for HOME, OVERVIEW, DOCUMENTATION, DOWNLOADS, SUPPORT, COMMUNITY, WIKI, and ABOUT US. A sidebar on the left lists categories: PROTEGE SOFTWARE, PLUG-INS, and ONTOLOGIES. The main content area is titled "download protégé" and lists three options: WebProtégé, Protégé Desktop 4.3, and Protégé Desktop 3.5.



The second screenshot shows a web browser window with the address bar containing `protege.stanford.edu/download/protege/3.5/installanywhere/Web_Installers/`. The page displays the Protege logo and the text "Protege 3.5 (build 663)". Below this, it states "Recommended Installation for Your Platform:" and provides a button for "Download Installer for Windows..." with a checked checkbox for "Include a VM in download". At the bottom, it reads "Installer created with InstallAnywhere. Copyright © 2012 Flexera Software LLC. All Rights Reserved."

Available Installers



Una vegada descarregat l'instal·lador procedim a la instal·lació amb la configuració per defecte del programa pel que polem el botó Next fins a finalitzar la instal·lació:

