



“Anàlisi actual de l’etiquetatge semàntic dels recursos: Schema.org i OpenGraphProtocol”

Daniel Oriol Villanueva Terrado
Màster Universitari en Enginyeria Informàtica

Sr. Felipe Geva Urbano

17/06/2016



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

FITXA DEL TREBALL FINAL

Títol del treball:	<i>Anàlisi actual de l'etiquetatge semàntic dels recursos: Schema.org i OpenGraphProtocol</i>
Nom de l'autor:	<i>Daniel Oriol Villanueva Terrado</i>
Nom del consultor:	<i>Felipe Geva Urbano</i>
Data de lliurament (mm/aaaa):	<i>06/2016</i>
Àrea del Treball Final:	<i>Web Semàntica i representació del coneixement</i>
Titulació:	<i>Màster Universitari en Enginyeria Informàtica</i>

Resum del Treball (màxim 250 paraules):

Al realitzar una cerca per Internet, sempre ens hem fet la mateixa pregunta: "Com s'estructura la informació dins la web?". Quan ens la fem, és perquè hem cercat informació específica i, com a resultat, els cercadors ens mostren molts resultats però no tots contenen informació íntegra a la pregunta formulada.

Arribats en aquest punt, entraria la web semàntica, la qual, donaria l'estructura necessària a la informació, per poder garantir una ràpida resposta des de qualsevol cercador i una resposta "real" dins dels paràmetres demanats per l'usuari. És a dir, que si l'usuari vol cercar "diferents models de televisors PHILIPS", li surtin –només– resultats de pàgines webs que tinguin aquesta informació, "models de televisors PHILIPS". Aquest etiquetatge, afavorirà en l'estructura de la informació, la seva pertinent assimilació dins del sistema, però també, en que les cerques siguin més ràpides i fiables 100%.

Actualment, hi ha força controvèrsia per saber quina ontologia és la més adequada.

En aquest TFM, ens centrarem en analitzar, a nivell funcional, els dos models estàndards d'etiquetatge de diferents dominis, l'Schema.org i TheOpenGraphProtocol.

Per poder realitzar aquest TFM, a part de la cerca i anàlisi d'informació, haurem d'analitzar diferents microdades, microformats i RDFas per trobar similituds i diferències entre aquests dos grans models d'etiquetatge.

Abstract (in English, 250 words or less):

We always wonder the same when we make a research at Internet: “How the information is structured on the Web ?” That is because we search some specific information and , as a result , web searches show us a lot of results but not all of them contain all the information we want to know.

At this moment, semantic web would start to give the necessary structure to the information, to ensure a fast answer from any searcher and a real answer inside the settings asked by the user. For example, if the user wants to find “different models of Philips televisions” he only find the results of web pages which has this information “Philip televisions”. This labeling helps in information structure, its assimilation inside the system, but in the search in order to be faster and reliable at 100%.

Currently, there is controversy to know which is the best ontology. In this project we are going to analyze , in a functional level , both standard models of labeling in different domains , the Schema.org and TheOpenGraphProtocol.

In order to realize this project, apart from information searching and analysis, we are going to analyze different microdata, microformats and RDFas to find similarities and differences between these labeling models.

Paraules clau (entre 4 i 8):

Semàntica, Ontologia, Schema.org, TheOpenGraphProtocol, RDFa, Microformat, Microdades

Índex de Continguts

1. Introducció.....	4
1.1 Context i justificació del Treball.....	4
1.2 Objectius del Treball.....	5
1.3 Enfocament i mètode seguit.....	5
1.4 Planificació del Treball.....	6
1.5 Breu sumari de productes obtinguts.....	6
1.6 Breu descripció dels altres capítols de la memòria.....	7
2. Resta de capítols.....	9
2.1 Definir: Microdades, Microformats i RDFas.....	9
2.2 Esquemes d' estructura del marcatge de dades.....	14
2.3 Anàlisi de webs representatives de cada domini.....	21
2.4 Anàlisi de les Microdades, Microformats i RDFas.....	40
2.5 Estudis amb Microdades, Microformats i RDFas, en.....	46
Schema.org i OGP	
2.6 Conclusions.....	55
Bibliografia.....	57

Índex de figures

Fig. 1 - Exemple de codificació en Microdades.....	10
Fig. 2 - Structured Data Testing Tool	11
Fig. 3 - Exemple de codificació en Microformat.....	12
Fig. 4 - Exemple de codificació en RDFa	12
Fig. 5 - Logo de Schema.org.....	14
Fig. 6 - Schema: Etiqueta Itemscope.....	15
Fig. 7- Schema: Etiqueta Itemtype.....	15
Fig. 8 - Schema: Etiqueta Itemprop.....	16
Fig. 9 - Schema: Categoria Thing.....	16
Fig. 10 - Schema: Estructura de marcatge.....	17
Fig. 11 - Schema: Categoria DataType.....	18
Fig.12- TOGP: Família NoVertical/Website.....	20
Fig.13- Schema: depuració de l'objecte "Place".....	23
Fig.14- Schema: depuració de l'objecte "Person".....	28
Fig.15- Schema: depuració de l'objecte "Organization".....	33
Fig.16- OGP: Etiquetatge HTML, propietat "og:title".....	34
Fig.17- OGP: Etiquetatge HTML, propietat "og:description".....	35
Fig.18- OGP: Etiquetatge HTML, propietat "og:img".....	35
Fig.19- OGP: Etiquetatge HTML, propietat "og:site name" i "og:locale".....	35
Fig.20- OGP: Depurador d'etiquetes OGP, per Facebook.....	35
Fig.21- Pàgina web del "Restaurant Nou Casa José".....	36
Fig.22- Depuració de la web "Casa José", per l'eina "Facebook for Developers"(1).....	38
Fig.23- Depuració de la web "Casa José", per l'eina "Facebook for Developers"(2).....	39
Fig.24: Exemple senzill d'una codificació i estructuració d'una pàgina web en HTML i amb microdades.....	41
Fig.25: Execució del codi HTML amb microdades.....	41
Fig.26: Exemple senzill d'una codificació i estructuració d'una pàgina web en HTML i amb microformats.....	42
Fig.27: Testeig d'una pàgina web en HTML amb l'eina "Structured Data Testing Tool", de Google.....	43
Fig.28: Execució del codi HTML amb microformats.....	43
Fig.29: Exemple complex d'una codificació en RDFa.....	44
Fig.30: Execució del codi HTML amb RDFa.....	45
Fig.31: Exemple d'inserció d'Schema i OGP en la mateixa web.....	46
Fig.32- Rastreig comú realitzat al Desembre del 2014 per webdatacommons: Totals.....	48
Fig.33- Rastreig comú realitzat al Desembre del 2014 per webdatacommons: Resultats per Format.....	49
Fig.34- Rastreig comú realitzat al Desembre del 2014 per webdatacommons: Grafiques(l).....	50

Fig.35- Rastreig comú realitzat al Desembre del 2014 per webdatacommons:	
Grafiques(II).....	50
Fig.36- Rastreig comú realitzat al Novembre del 2015 per webdatacommons: Totals.....	51
Fig.37- Rastreig comú realitzat al Novembre del 2015 per webdatacommons:	
Ressultats per Format.....	52
Fig.38- Rastreig comú realitzat al Novembre del 2015 per webdatacommons:	
Grafiques(I).....	53
Fig.39- Rastreig comú realitzat al Novembre del 2015 per webdatacommons:	
Grafiques(II).....	53

1. Introducció

1.1 Context i justificació del Treball

Avui en dia, la semàntica i estructuració de les pàgines web, no porta cap mena de control. Depèn dels propis usuaris de la xarxa, i dels seus coneixements en la matèria, ja que utilitzen el marcat d'etiquetatge que creuen oportú però, no hi ha cap "normativa" i/o "protocol", alhora de crear una pàgina web i etiquetar l'informació emmagatzemada dins d'ella...

És necessari saber com guarden, classifiquen i identifiquen l'informació les pròpies webs. Si tinguéssim clar quina metodologia utilitzen, podríem obtenir més bons resultats en els propis cercadors i, millorar en l'obtenció d'informació específica.

Tenint coneixença dels formats d'etiquetatge i, les pertinents estructures de marcatge de dades, es vol realitzar una cerca per Internet, per saber:

- Com s'utilitzen?
- Quina és la metodologia més utilitzada? Perquè?
- Quins beneficis en trèiem?
- Quines diferències hi ha entre un marcatge i l'altre?
- És realment necessari?

Veient aquestes preguntes, podem adonar-nos de l'importància que té aquest estudi i de les millores que pot arribar a originar, en el camp de la semàntica. Sabem que, avui en dia, no es sap si es tendirà amb un determinat marcatge o l'altre... però, es possible, que facilitem informació necessària per determinar cap a on va el futur de la semàntica...

Per poder aportar les nostres afirmacions, s'haurà de cercar en moltes pàgines webs, mirar les seves estructures, etiquetar-les, estudiar-les, crear webs amb diferents estructures i provar-les. Totes aquestes proves, més l'obtenció d'informació de les noves tecnologies i llenguatges, faran que tinguem clars les diferències entre elles (llenguatges, etiquetatges, evolució, quantitat de gent que l'utilitza, etc) i que ens puguem "decantar" en un costat (Schema), on en un altre (OGP).

1.2 Objectius del Treball

Durant el desenvolupament del TFM, ens centrarem en la web semàntica, en les estructures de dades i els formats que presenten les diferents webs d'Internet.

Per poder-ho fer, ens centrarem, en:

- Definir la web semàntica.
- Definir l'Ontologia.
- Definir els models estàndards d'etiquetatge; "Schema.org" i "TheOpenGraphProtocol".
- Utilitzar els llenguatges següents; "Microdades", "Microformats" i "RDFas", en la representació d'ontologies.
- Realitzar l'anàlisi de dominis, de; "Schema.org" i "TheOpenGraphProtocol".
- Realitzar proves amb diferents formats.
- Extreure conclusions dels dos estàndards d'etiquetatge.

1.3 Enfocament i mètode seguit

L'estratègia que seguirem, serà primer que tot, informar-nos del què és una web semàntica i què és una ontologia. Un cop tinguem clar aquests conceptes, tindrem que saber què significa els estàndards d'etiquetatge, i quins hi ha de principals.

Al saber en quin context ens movem haurem de determinar quins llenguatges utilitzarem per poder utilitzar les respectives ontologies. En aquest cas, s'utilitzaran les Microdades, els Microformats i els RDFas.

Per poder realitzar una visió del problema, farem cerques –en el cercador Google- de diferents temes. Un cop tinguem els resultats obtinguts, ens dedicarem a identificar les diverses estructures de cada pàgina web. En fer-ho, podrem determinar quin marcatge és el més utilitzat dels dos (el nombre de proves realitzades serà força elevada per poder estar "una mica" segurs del resultat obtingut).

A partir d'aquí, es començarà a fer proves (es crearan webs amb les diferents estructures) per poder veure si, efectivament, aquest tipus d'estructura és òptima per l'etiquetatge massiu de les pàgines webs i, a posteriori, per poder tornar resultats a partir de la seva estructura...

En la meua humil opinió, crec que és una bona metodologia a utilitzar per poder anar avançant "poc a poc" i correctament. Al no ser un projecte a on es "crea" cap hw ni sw, és normal que hi hagi moltes deduccions que, alhora arribaran a unes conclusions. Per poder haver les deduccions, han d'haver un seguit de "proves" (realitzades a internet i documentades a la memòria), les quals mostraran el perquè s'arriba a aquestes deduccions i/o conclusions. També, ens basarem en uns estudis, realitzats per WebDataCommons, que van realitzar durant el mes de Desembre del 2014 i el mes de Novembre del 2015. Al ser relativament actuals, podrem basar-nos en els seus resultats per reafirmar les nostres pròpies conclusions.

1.4 Planificació del Treball

NOM	DURADA	INICI	FINAL
PAC1	14 dies	25/02/2016	09/03/2016
Taula de Fites	14 dies	25/02/2016	09/03/2016
Diagrama de Gantt			
PAC2	42 dies	10/03/2016	20/04/2016
Anàlisi i disseny del TFM	42 dies	10/03/2016	20/04/2016
Anàlisi de requisits	5 dies	10/03/2016	14/04/2016
1. Definir: Microdades, Microformats i RDFas. <ul style="list-style-type: none"> Per a què serveixen? Per a què les utilitzem? Per a què s'ha de marcar la informació dins la web? 	18 dies	15/03/2016	01/04/2016
2. Esquemes d'estructura del marcatge de dades: <ul style="list-style-type: none"> En Schema.org En TheOpenGraphProtocol(OGP) Diferències i similituds 	19 dies	02/04/2016	20/04/2016
PAC3	21 dies	21/04/2016	11/05/2016
3. Anàlisi de webs representatives de cada domini, en: <ul style="list-style-type: none"> Schema.org TheOpenGraphProtocol (OGP) 	10 dies	21/04/2016	30/04/2016
	5 dies	21/04/2016	25/04/2016
	5 dies	26/04/2016	30/04/2016
4. Anàlisi de les Microdades, Microformats i RDFas: <ul style="list-style-type: none"> Estructura Comportament 	11 dies	01/05/2016	11/05/2016
	6 dies	01/05/2016	06/05/2016
	5 dies	07/05/2016	11/05/2016
PAC4	37 dies	12/05/2016	17/06/2016
5. Estudis amb microdades, microformats i RDFas, en Schema.org i OGP: <ul style="list-style-type: none"> Simbiosi entre els diferents vocabularis(Schema,ogp) i les diverses sintaxis (microdades, microformats i RDFas) Estudis realitzats per WebDataCommons 	15 dies	12/05/2016	26/05/2016
6. Conclusions	15 dies	26/05/2016	09/06/2016
7. Memòria: <ul style="list-style-type: none"> Redacció i correcció de la memòria Presentació 	114 dies	25/02/2016	17/06/2016
	114 dies	25/02/2016	17/06/2016
	5 dies	13/06/2016	17/06/2016

1.5 Breu sumari de productes obtinguts

El producte serà la memòria realitzada amb tota la informació obtinguda. En diferents cercadors i navegadors, s'han realitzat les diferents proves de webs bàsiques que, gràcies a aquestes, hem pogut arribar a obtenir una sèrie de conclusions referents al tema que estem tractant...

Per tant, el producte obtingut és la informació extreta a partir dels resultats que hem extret de proves i cerques d'Internet. Aquestes proves i conclusions,

ens han acompanyat al llarg del camí per poder determinar; Cóm son?, Cóm es troben estructurades?, Cóm estan etiquetades?, Cóm facilita l'etiquetatge amb la cerca?, Quins llenguatges utilitzen?, son òptims?, perquè?, etc...

1.6 Breu descripció dels altres capítols de la memòria

1. Definició, de; Microdades, Microformats i RDFas.

En aquest capítol, ens centrarem en analitzar les tres maneres de marcatge semàntic que té la pròpia web. Per poder-ho fer, ens basarem en tres preguntes bàsiques; Per a què serveixen?, Per a què les utilitzem?, Per a què s'ha de marcar la informació dins d'una web?. Totes aquestes preguntes, al respondre-les, obtenim l'essència del significat de cadascuna d'elles.

2. Esquemes d'estructura del marcatge de dades.

Es tracta d'explicar la metodologia que s'utilitza en marcar la informació dins de la pròpia web. Un cop assolida aquesta informació, es genera una explicació detallada de cada estructura i de cada etiqueta que forma part tant en "*Schema.org*", com en "*TheOpenGraphProtocol*".

3. Anàlisi de webs representatives de cada domini, en: Schema.org i TheOpenGraphProtocol.

Per poder treure profit de la web semàntica, son precisos els llenguatges semàntics més potents. Aquests, son els llenguatges de marcatge capaços de representar el coneixement basant-se en l'ús de les metadates i ontologies. No obstant, uns altres es converteixen en un objecte ric dins d'una xarxa social. Aquesta peculiaritat, permet definir cóm es vol que es publiqui el contingut de la web, dins d'una xarxa social. Les dos funcionen, respectivament be dins dels seus camps. Ara identificarem, el perquè funcionen tant be i com ho fan,...a part de desglossar la seva estructura interna.

4. Anàlisi de les Microdades, Microformats i RDFas.

S'intenta diferenciar els tres llenguatges que donen la sintaxis per poder estructurar les dades i tenir-les com a vocabulari propi o com a una barreja de diversos vocabularis. El objectiu és el mateix però, s'executen de diferents maneres. En aquest capítol, desglossem cadascun d'ells i els relacionem amb exemples creats per nosaltres (testejats prèviament a la xarxa).

5. Estudis amb microdades, microformats i RDFas, en Schema.org i OGP:

La primera part es basa en relacionar els diferents vocabularis amb les seves respectives sintaxis i comentar els trets característics de cada una. També, es realitza una reflexió de la importància d'insertar més d'un vocabulari dins de la pròpia web.

En la segona part, ens centrem en uns estudis de WebDataCommons els quals, ens facilita molta informació de com es troben estructurades les webs avui en dia (període del Desembre del 2014 fins al Novembre del 2015). Partint dels resultats obtinguts, podem establir una evolució i regressió de les diferents sintaxis treballades, confirmant tot el que s'ha dit al llarg del TFM.

6. Conclusions.

L' objectiu és poder arribar a unes conclusions sobre els dos tipus de marcatges. Realitzar una valoració amb tots els aspectes a tenir en compte (evolució, software, sintaxis, compatibilitat en navegadors i programes,etc) i poder determinar quin és millor o, quin evolucionarà més favorablement...

2. Resta de capítols

2.1 Definir: Microdades, Microformats i RDFa

- a. Per a què serveixen?
- b. Per a què les utilitzem?
- c. Per a què s' ha de marcar la informació dins la web?

Fins fa poc, les pàgines web estaven estructurades i etiquetades de moltes maneres. Aquest factor, influenciava negativament alhora de realitzar cerques d'informació, en diferents cercadors. Per aquesta raó, es va decidir que les pàgines web havien d'estar estructurades i etiquetades en els seus continguts. En aquell moment, es van donar a conèixer "**Schema.org**" i el "**OpenGraphProtocol**", entre d'altres. La funció d'aquests dos és la mateixa, donar format i significat al context de la pròpia pàgina web. Per poder-ho garantir, es basa en, els microformats.

Les microdades, microformats i els RDFa, són tres classes de marcat semàntic. Son diferents maneres d'etiquetar els continguts per a que, el cercador, compregui millor l' informació a cercar.

A diferència de les microdades, creades en llenguatge HTML5, agrupa els continguts amb l'estructura *nom-valor*, els microformats i els RDFa, estan més pensats per XHTML. Els microformats, utilitzen generalment l'atribut "class", per estructurar l' informació. Mentre, els RDFa, son un conjunt d'extensions XHTML, que aprofiten i generalitzen atributs simples.

Microdades (*rich snippets*):

Les microdades, son les propietats que componen l'etiquetatge de l'HTML5. Aquestes propietats, ens permeten crear el contingut de les etiquetes, que en essència és la part semàntica de cada una d'elles....Sense les microdades, un cercador (p.ex. *Google*), no aconseguiria relacionar, amb exactitud, el significat de cada una de les paraules que es volguessin cercar. Amb les microdades, el text a cercar es molt més entenedor i comprensible, alhora de realitzar la cerca. Arribat en aquest punt, el cercador, abans de realitzar la cerca, entén que:

- **Les dades estan definides:** Entén el que ha de cercar. Reconeix el camp d'informació al que ha de cercar l' informació demanada.
- **Les dades es troben estructurades:** Per poder entendre el significat de la cerca, la informació ha d'estar classificada per la seva pròpia comprensió Com més estructurat, més comprensible pel cercador....
- **Les microdades han d'estar relacionades entre si:** El cercador ha de buscar informació de múltiples paraules, aquestes separades per espais, les quals han de compartir alguna característica en comú. Si no fos així, el cercador no disposaria de filtres específics i, el resultat de la cerca, seria molt genèric...

Una característica important de les microdades, son les incrustacions (*embedded items*) que poden fer incrustacions de microdades dins d'altres. Aquest factor, origina que a part de tenir esquemes declarats, puguem utilitzar-les i crear-ne de noves, només afegint incrustacions.

Google ha recomanat l' utilització de la web semàntica, mitjançant les microdades. També ha creat una eina online(**Google Webmaster Tools for Rich Snippets**) per testejar la utilització de les dades estructurades dins de les pàgines webs. Aquesta eina es troba on-line i es pot utilitzar gratuïtament.

Les microdades, utilitzen atributs senzills en les etiquetes HTML(generalment, o <div>), per assignar noms curts i descriptius als elements i a les propietats.

Hi ha diferents tipus de microdades:

- **Itemscope:** Indicador de que l'element especificat, serà el contenidor d'informació important. És del tipus booleà.
- **Itemtype:** Indica quin tipus d'informació ens trobarem, diccionari, quin vocabulari s'utilitzarà. Indica una direcció URL, on especifica aquesta informació.
- **Itemid:** Atribut que dona informació sobre un element comú a tot Internet. Es com si creéssim una fitxa que contingui tota la informació que vulguem donar.
- **Itemref:** Es dona quan volem fer referència a varis elements del mateix contingut. Se'ls i assigna un nom a cada un.
- **Itemprop:** Es la propietat i el valor que volem mostrar.

```
<div itemscope itemtype="http://data-vocabulary.org/Person">  
Soc <span itemprop="name">Daniel O. Villanueva Terrado.</span>  
Pero, també em diuen <span itemprop="nickname">Dani.</span>.  
La nostra pàgina web és:  
<a href="http://www.ilerna.com" itemprop="url">www.ilerna.com</a>  
Treballo en el <span itemprop="title">  
Departament Informàtica,</span> com a <span itemprop="affiliation">professor.</span>.</div>
```

[Fig.1 - Exemple de codificació en Microdades.](#)

En definitiva, les microdades estan recomanades per **Google** i també per **HTML5** i **W3C**. També, si es vol aprofitar de les avantatges d'estar en les posicions de d'alt de tot dels "resultats de cerca", de Google. Has d'inserir les microdades a la teva pàgina web.

Microformats:

Els microformats, són la manera més elegant d'enriquir i semantitzar el seu contingut. Aquesta especificació, del propi contingut semàntic, afavoreix a una bona comprensió del codi per a les pròpies màquines, alhora de tractar la informació requerida.

Cal esmentar que, els microformats, són la part de codi HTML (XHTML), que s'utilitza per poder; categoritzar, donar significat a un o més objectes, ajudar als cercadors en la seva comprensió i assimilació dels resultats de la cerca, etc....Tot això, s'aconsegueix mitjançant la utilització de les classes de les etiquetes "divs" i "spans" per poder marcar-los. Aquesta, és la manera més comú d'utilitzar els microformats.

Una manera fiable de saber si ho has fet be, és anant a l'eina de Google, "**Structured Data Testing Tool**", per comprovar si,efectivament, l'has creat correctament.

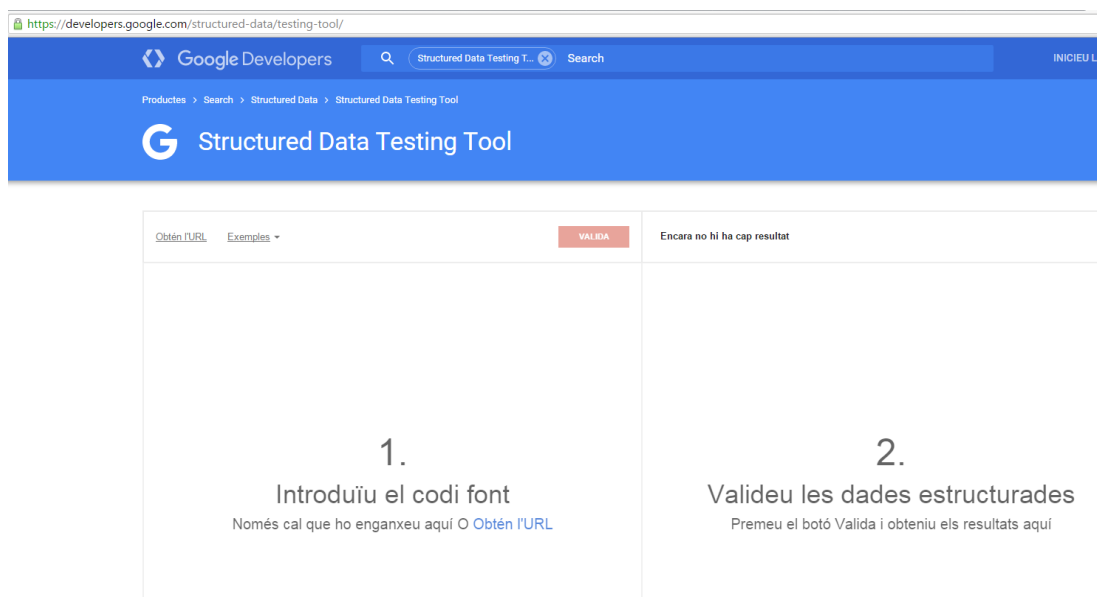


Fig.2 – Structured Data Testing Tool.

Url: <https://developers.google.com/structured-data/testing-tool/>

Tipus de microformats:

- **hAtom:** Utilitzat per marcar feed RSS atoms.
- **hCalendar:** Per poder senyalitzar events.
- **hCard:** És una tarja de contacte per a marcar les dades personals i/o empresarials.
- **adr:** Marcat per a direccions físiques.
- **geo:** Geolocalització de llocs per longitud i latitud.
- **hReview:** Opinions d'un producte, web, article, etc.
- **hResume:** Currículums.
- **XFN:** Enllaços per relacions socials.
- **XOXO:** Esquematització del contingut, mitjançant llistes.

```

]<div class="vcard">Soc 
<strong>Daniel O. Villanueva Terrado.</strong>
<span>Centre Ilerna</span>de
<span class="org">Departament Informàtica.</span>
<span>Carrer Pau Casals.</span>
<span>Lleida.</span>, <span>Catalunya.</span>
<span>25001</span>
</div>

```

Fig.3 - Exemple de codificació en Microformat.

RDFa:

És una manera d'etiquetar, el contingut, per descriure un tipus específic d'informació. Per exemple, la opinió sobre un restaurant, un concert, una persona o una llista de productes. Tots aquests tipus d'informació, s'anomenen entitats o elements. Cada entitat, disposa d'una sèrie de propietats. Sovint, el RDFa el que fa es, utilitzar atributs simples en etiquetes XHTML, per poder assignar noms curts i descriptius a les pròpies entitats i a les seves propietats . Per poder-ho fer, utilitza l'etiqueta <Body> (dins del <head>, on s'utilitza l' informació meta). Aquesta, és la que completa l' informació que es mostra per pantalla i l'usuari l'interpreta. També, aquesta informació, aporta riquesa semàntica, la qual millora el posicionament de la web respecte les demés, ja que, li estem donant més dades i més estructura per poder categoritzar i indexar l' informació de manera més efectiva.

```

]<div xmlns:v="http://rdf.data-vocabulary.org/#" typeof="v:Person"> Soc <span
property="v:name">Daniel Oriol Villanueva Terrado.</span>. Treballo al<span
property="v:title">Centre Ilerna</span> de <span
property="v:affiliation">Professor</span>. La nostra pàgina web és: <a
href="http://www.ilerna.com/" rel="v:url">www.ilerna.com</a>.

```

Fig.4 - Exemple de codificació en RDFa.

Diferència entre les microdades, els microformats i els RDFas:

	Microdades	Microformats	RDFa
Model de dades:	Arbre	Arbre	Graf
Complexitat del domini de la sintaxi:	Mig	Baix	Alt
Dominis:	Genèrics	Específics	Genèrics
Parsers:	Genèrics	Específics	Genèrics
Desenvolupament:	Descentralitzat	Centralitzat	Descentralitzat
Combinació de vocabularis	Es combinen	Difícil combinar	Es combinen
Amplitud del Vocabulari:	Amplia	Reduïda	Il·limitada
Llenguatge utilitzat:	HTML5	HTML / XHTML	XHTML
Model d'etiquetatge:	 / <div> Utilitzen ítems(nom-valor): itemscope, itemtype, itemid, itemprop, itemref	Reutilitza els següents atributs: Class, rel, rev, title	 / <div> Utilitza els següents atributs: about, datatype, inlist, prefix, property, resource, typeof, vocab

Els tres sistemes tracten de resoldre el mateix problema, però cadascun, de manera diferent. Això origina diferències entre ells. Sembla evident que, el més còmode i senzill és la microdada. Aquest fet, es dona per la possibilitat d'agrupar totes les etiquetes semàntiques i diferenciar-les de la resta. També, comptar amb l'ajut d' Schema.org, origina tenir un cert nivell de normalització en l'etiqueta i, com a segona característica, una major flexibilitat en la descripció (possibilitat de combinar diversos esquemes per poder crear un altra de més complexa i precisa).

Entre els tres tipus de marcat, Google es decanta per la utilització de les microdades. Aquest, considera que aquest format és més còmode, net i que ofereix l'equilibri entre una possible ampliació de RDFa i la senzillesa utilització dels microformats.

L' utilització del etiquetatge semàntic, permet una millor estructuració de l' informació i una profunda comprensió del significat, per part del cercador. També, un factor a tenir en compte es, que el seu ús incrementa fins a un 30%, la visibilitat de la pròpia pàgina web.

2.2 Esquemes d'estructura del marcatge de dades

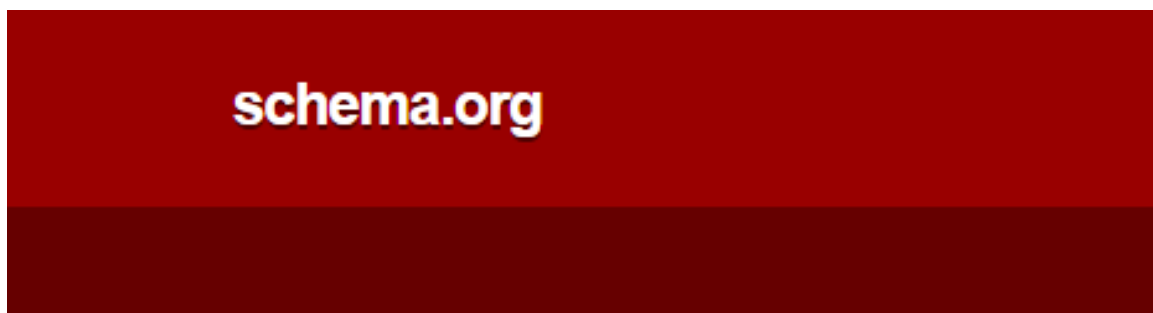
El marcatge de dades, o llenguatge de marques, és un llenguatge informàtic per poder codificar un document que incorpora etiquetes i/o marques, les quals, donen informació complementària sobre l'estructura del text o sobre la seva manera de mostrar-ho...

Com a llenguatge, per excel·lència, tenim l'HTML (és un llenguatge de marcatge d'hipertext). Els llenguatges de marcatge, es poden confondre amb els llenguatges de programació. Val a dir que, hi ha una gran diferència entre ells, els llenguatges de marcatge, no tenen funcions aritmètiques o variables....

En els llenguatges de marcatge, es diferencien tres tipologies, encara que es poden combinar diverses classes en un document:

- **Marcatge de Presentació:** Indica el format que tindrà el text. S'utilitza per configurar la presentació d'un document. Resulta molt fàcil d'utilitzar-lo, si comptem amb poca informació. És molt complicat portar un manteniment, o la seva pertinent modificació, si comptem amb un volum d'informació considerable....
- **Marcatge de procediments:** També està dirigit a la presentació del document. Per mostrar la informació, s'ha d'executar el codi, en el mateix ordre en que està estipulat que aparegui.
- **Marcatge descriptiu (o semàntic):** Utilitza etiquetes per poder donar format als diferents fragments de text. No especifiquen com s'han de mostrar o de quina manera han d'esser mostrats. Les etiquetes poden contenir qualsevol classe de metadades. Un factor a tenir en compte seria la seva flexibilitat, ja que, els fragments de text s'etiqueten *tal i com son*, i no *tal i com han d'anar*. També, simplifica la tasca de canviar totalment el format d'un text, ja que aquesta informació, està separada del propi text. Tanmateix, el marcatge descriptiu, al realitzar-se amb etiquetes, facilita molt la feina d'edició i manteniment al haver-hi les etiquetes que agrupen l'informació i es pot manipular globalment.

En aquest punt, analitzarem les diferents tipologies de marcatge que ens trobem en els recursos Schema.org i OpenGraphProtocol:



[Fig.5- Logo Schema.org](#)

Schema.org:

Tenint en compte que, Schema, és l'únic sistema de marcatge de dades, compatible amb els principals motors de cerca (Google, Yahoo, Bing, etc.).

- **Les etiquetes, per poder marcar el contingut, son:**
 - **Itemscope:** Es per definir un determinat objecte. Amb aquesta etiqueta, indiquem als diferents cercadors que, a continuació, hi ha informació amb microdades.

```
<div itemscope>  
  
<h1>UOC</h1>  
  
<span>TFM-Web semantica i Representacio del Coneixement.</span>  
  
<span>Treball Final de Master</span>  
  
<a href="http://www.uoc.edu/portal/ca/index.html">Acces al Campus.</a>  
  
</div>
```

[Fig.6 – Schema: Etiqueta Itemscope](#)

- **Itemtype:** Indica el format, el tipus de dades, les quals estem marcant. Es a dir, serveix per indicar als cercadors de quin tipus de dades estem tractant.

```
<div itemscope itemtype = "http://schema.org/CollegeOrUniversity">  
  
<h1>UOC</h1>  
  
<span>TFM-Web semantica i Representacio del Coneixement.</span>  
  
<span>Treball Final de Master</span>  
  
<a href="http://www.uoc.edu/portal/ca/index.html">Acces al Campus.</a>  
  
</div>
```

[Fig.7- Schema: Etiqueta Itemtype](#)

- **Itemprop:** Aquesta, indica la propietat de la família que hem definit el text.

```
<div itemscope itemtype = "http://schema.org/CollegeOrUniversity">
  <h1 itemprop="name">UOC</h1>
  <span>TFM: <span itemprop="department">Web semantica i Representacio del Coneixement.</span>
  <span>Treball Final de Master</span>
  <a href="http://www.uoc.edu/portal/ca/index.html" itemprop="url">Acces al Campus.</a>
</div>
```

Fig.8- Schema: Etiqueta Itemprop

En les tres imatges anteriors, he realitzat un marcatge, dins de la família “**Organization/CollegeOrUniversity**”, etiquetant; **name** (nom del Centre), **department** (he insertat el nom del projecte, com si fos un Departament del propi Centre) i **url** (l’enllaç amb el portal del Centre). Gràcies a aquest marcatge, podem estar segurs que, si ens busquessin, els cercadors sabrien exactament, el contingut de la nostra web....

Si ens fixem en Schema.org, l’informació que ens facilita en la seva web (schema.org/docs/full.html), trobem que es basa en dos categories:

- **Thing**: Una categoria, amb una llarga llista de continguts molt amplis, per poder definir i complimentar qualsevol “cosa”(**Thing**).

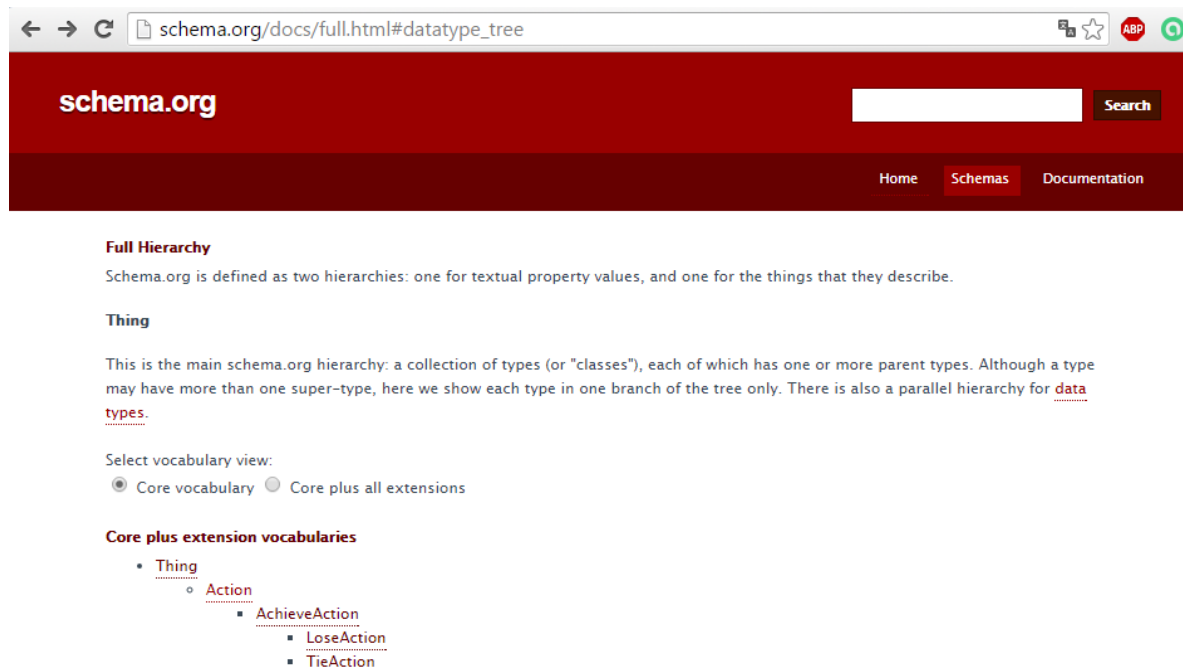


Fig.9- Schema: Categoria Thing

Si anem a la web oficial d' "Schema.org" (schema.org/docs/full.html), podem obtenir l'esquema de l'estructura que segueix, en el marcatge de dades.

Schema.org

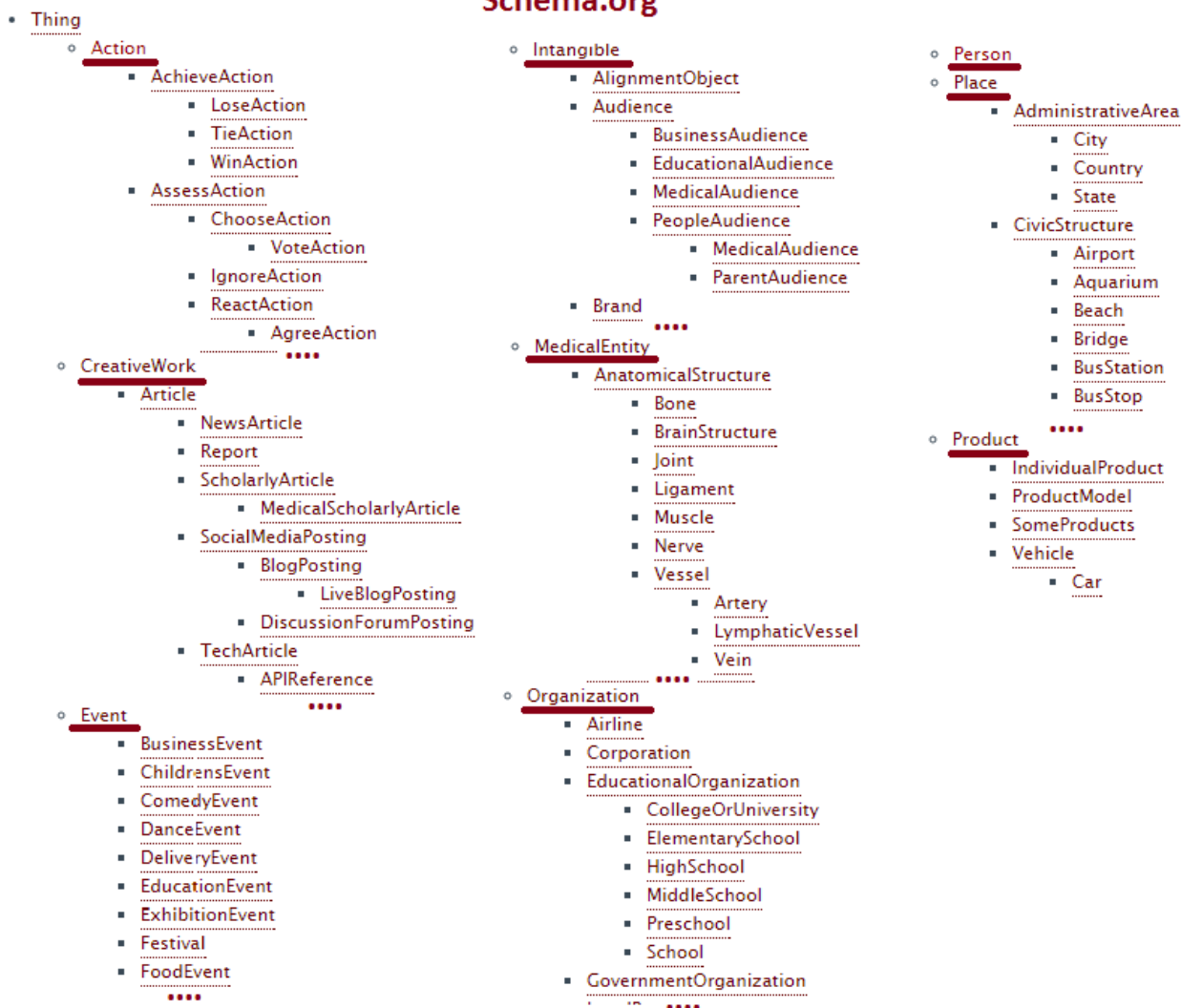


Fig.10- Schema: Estructura de marcatge

- **Data Type:** Categoria que facilita la definició i complementació dels diferents "tipus de dades" (**Data Type**).

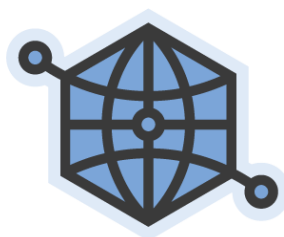
Data Types

- DataType
 - Boolean
 - False
 - True
 - Date
 - DateTime
 - Number
 - Float
 - Integer
 - Text
 - URL
 - Time

Fig.11- Schema: Categoria DataType

Aquestes dos últimes categories, complimenten l'etiquetatge que hem vist abans (*itemscope*, *itemtype* i *itemprop*). Son els apartats en què es constitueixen cada un dels objectes que s'han de definir per a poder ser compresos pels diferents cercadors (sobretot, per Google). Sense aquests, els cercadors no entendrien la informació a cercar, i molt menys a trobar....

The Open Graph protocol



The Open Graph Protocol:

Aquest protocol, va ser creat per Facebook. Compta amb la documentació que es troba, en la pàgina oficial (<http://opengraphprotocol.org/>). És un mètode senzill, el qual, ens permet incloure meta informació dins la pàgina web, i d'aquesta manera, convertir-la en un Objecte Social. Un cop definit l'objecte social, aquest, pot interactuar amb d'altres objectes socials (Social Graph), com; share de Google+, like de Facebook, etc....

Per a que el protocol s'utilitzi en una web, s'han d'introduir una sèrie d'etiquetes dins del codi HTML, entre les etiquetes **<head></head>**, depenen del contingut que es vulgui publicar i tenint en compte els diferents formats que s'han de mostrar (imatges, enllaços, etc.). De la llarga sèrie de propietats, que ens dona el protocol, només tres son indispensables:

- **og:type:** Seria el tipus d'objecte. Aquests, s'agrupen en les verticals. Cada una d'elles, disposa del seu conjunt de noms. Els valors que té og:type, sempre tenen un prefix. Aquest, facilita la comprensió amb els tipus d'espais de noms definits per l'usuari, que sempre tenen dos

punts.

Els tipus que té predefinits, com a “**Object Types**”, son:

- **Music:**
 - Namespace URI: <http://ogp.me/ns/music#>
 - Tipus de valors:
 - **music.song.**
 - **music.album.**
 - **music.playlist.**
 - **music.radio_station.**
- **Video:**
 - Namespace URI: <http://ogp.me/ns/video#>
 - Tipus de valors:
 - **video.movie:**
 - video:actor.
 - video:actor:role.
 - video:director.
 - video:writer.
 - video:duration.
 - video:release_date.
 - video:tag.
 - **video.episode:**
 - video:actor.
 - video:actor:role.
 - video:director.
 - video:writer.
 - video:duration.
 - video:release_date.
 - video:tag.
 - video:series.
 - **video.tv_show.**
 - **video.other.**
- **No Vertical:**
 - **Article - Namespace URI:** <http://ogp.me/ns/article#>
 - Tipus de valors:
 - article:published_time.
 - article:modified_time.
 - article:expiration_time.
 - article:author.
 - article:section.
 - article:tag.
 - **Book – Namespace URI:** <http://ogp.me/ns/book#>
 - Book:autor.
 - Book:isbn.
 - Book:release_date.
 - Book:tag.
 - **Profile – Namespace URI:**

<http://ogp.me/ns/profile#>

- Profile:first_name.
- Profile:last_name.
- Profile:username.
- Profile:gender.

○ **Website-Namespace URI:**

<http://ogp.me/ns/website#>

- **og:image:** És una URL que apuntarà a una imatge i, aquesta, representarà l'objecte definit.
- **og:url:** La pròpia *url* canònica s'utilitzarà com a identificatiu permanent del propi objecte.
- Les demés, s'utilitzaran a partir de les necessitats que tinguem alhora de crear-les.

```
<html prefix="og: http://ogp.me/ns#">
  <head>
    <title>UOC</title>
    <meta property="og:title" content="TFM" />
    <meta property="og:type" content="website" />
    <meta property="og:image"
      content="http://cv.uoc.es/estudiant/\_resources/img/novetats/home/uoc\_mobil.jpg" />
    <meta property="og:url" content="http://www.uoc.edu/portal/ca/index.html" />
  </head>
  ...
</html>
```

[Fig.12- TOGP: Família NoVertical/Website](#)

En la imatge anterior, he realitzat un marcatge, dins de la família **"NoVertical/Website"**, etiquetant; **og:title** (nom del *Treball Final de Màster*), **og:type** (he insertat el contingut del objecte, una *web*), **og:image** (l'enllaç de la imatge, del marcatge, amb la ubicació de la pròpia imatge a internet) i **og:url** (l'enllaç del propi objecte definit dins la xarxa). Gràcies a aquest marcatge, podem estar segurs que, si ens busquessin, els cercadors sabrien exactament, el contingut de la nostra web....

2.3 Anàlisi de webs representatives de cada domini:

- a. Schema.org.
- b. TheOpenGraphProtocol.

Schema.org

En Schema, el primer element que destaca –dins de l’etiqueta HTML- és l’ **itemscope**, el qual identifica el contingut de la pròpia web. Si volem identificar el tipus d’element, necessitarem l’etiqueta **Itemtype**. Aquesta, identifica qualsevol tipus d’element que es trobi definida dins de la jerarquia de tipus del portal Schema.org. En aquest portal hi han definits varis elements però, n’hi ha quatre propietats que son les més importants i utilitzades; *nom*, *descripció*, *imatge* i *url*. Les demés propietats comparteixen d’altres propietats, un clar exemple pot ser en els tipus “*Organization*”, “*LocalBusiness*”, etc.

Anem a analitzar diversos objectes d’Schema.org:

- “**Place**”(Lloc): Amb aquest objecte es pot representar una ubicació, un lloc, la localització d’un Restaurant, etc... A part, de tota la informació “extra” que vulguis afegir-hi a aquesta ubicació/lloc.

CODI:

```
<div itemscope itemtype="http://schema.org/Place">
  <span itemprop="address" itemscope itemtype="http://schema.org/PostalAddress">
    <span itemprop="addressLocality">La Seu Vella</span>
    <span itemprop="addressRegion">Lleida</span>
    <span itemprop="addressCountry" itemscope
itemtype="http://schema.org/Country">
      <span itemprop="name">Catalunya</span>
    </span>
  </span>
  <div itemprop="aggregateRating" itemscope
itemtype="http://schema.org/AggregateRating">
    <span itemprop="name">Calificacions:</span>
    <span itemprop="itemReviewed">Maig del 2016:</span>
    <span itemprop="ratingCount">123</span>
    <span itemprop="ratingValue">97</span>
    <span itemprop="worstRating">1</span>
    <span itemprop="bestRating">100</span>
  </div>
  <div itemprop="containedIn" itemscope itemtype="http://schema.org/Place">
    <span itemprop="name">Centro Historic de Lleida</span>
  </div>
  <div itemprop="event" itemscope itemtype="http://schema.org/Event">
    <span itemprop="name">11 de Maig: FESTA MAJOR </span>
  </div>
</div>
```

```

<div itemprop="location" itemscope itemtype="http://schema.org/Place">
  <span itemprop="name">Estacio de Lleida</span>
  <span itemprop="address">Davant de Rambla Ferran</span>
</div>
<span itemprop="startDate">2016-05-11:00:00</span>
<div itemprop="offers" itemscope itemtype="http://schema.org/Offer">
  <span itemprop="name">Ruta: Bus turistic / Visita panoramica</span>
  <span itemprop="price">8€</span>
  <span itemprop="url">http://www.turismedelleida.cat/informacio_practica/visitas-
guiades-1/grups-i-col-lectius/rutes-turistiques/ruta-bus-turitic</span>
</div>
</div>
<span itemprop="faxNumber">123-456-7890</span>
<div itemprop="geo" itemscope itemtype="http://schema.org/GeoCoordinates">
  <span itemprop="latitude">41,37,04 Nort </span>
  <span itemprop="longitude">0,37,36 Oest</span>
</div>
<span itemprop="globalLocationNumber">1234567890123</span>
<span
itemprop="hasMap">https://www.google.es/maps/place/La+Seu+Vella/@41.618084
6,0.6246873,17z/data=!3m1!4b1!4m5!3m4!1s0x12a6e068350330f9:0x8aca0fa5dd06
1a60!8m2!3d41.6180846!4d0.626876</span>
<span itemprop="interactionCount">248,689 usuarios</span>
<span itemprop="isicV4">pvjm</span>
<span
itemprop="logo">https://www.google.es/url?sa=i&rct=j&q=&esrc=s&source=imgres&
cd=&cad=rja&uact=8&ved=0ahUKEWjMg6alstLMAhWF6xoKHejQA-
MQjRwIBw&url=https%3A%2F%2Fen.wikipedia.org%2Fwiki%2FOld_Cathedral_of_L
leida&psig=AFQjCNFnPjWnolCcw4oWe-
5vLeuwuALbA&ust=1463068778021576</span>
<span itemprop="openingHoursSpecification">00:00:00</span>
<div itemprop="photo" itemscope itemtype="http://schema.org/ImageObject">
  <span itemprop="name">La_Seu_Vella.jpg</span>
</div>
<div itemprop="review" itemscope itemtype="http://schema.org/Review">
  <span itemprop="name">Ninguna</span>
  <div itemprop="author" itemscope itemtype="http://schema.org/Person">
    <span itemprop="givenName">Lleida</span>
  </div>
</div>
<span itemprop="telephone">123-456-7890</span>
<span itemprop="alternateName">Lerida</span>
<span itemprop="description">Ciutat emblematica de les terres de Ponent.</span>
<span
itemprop="image">https://www.google.es/url?sa=i&rct=j&q=&esrc=s&source=images
&cd=&cad=rja&uact=8&ved=0ahUKEwiqmYn2s9LMAhXLcBoKHYYHrB_kQjRwIBw&ur
l=http%3A%2F%2Fwww.bcnatfilmcommission.com%2Fes%2Fmunicipios%2Flleida
&psig=AFQjCNF7ORFbEEuMeB_Tic_hy_b371B6bw&ust=1463069264896404</spa
n>
<span itemprop="name">Malec&oacute;n</span>

```

```

<span itemprop="potentialAction">Descanso</span>
<span itemprop="sameAs">http://www.nh-hoteles.es/hotel/nh-lleida-
pirineos</span>
<span
itemprop="url">https://www.google.es/url?sa=t&rct=j&q=&esrc=s&source=web&cd=1
&cad=rja&uact=8&ved=0ahUKEWjO5lCrNLMAhVDChoKHUOGAmgQFggkMAA&url
=https%3A%2F%2Fes.wikipedia.org%2Fwiki%2FL%25C3%25A9rida&usg=AFQjCN
FA9wgivivok89z-95MCeXTm76FtSg</span>
</div>

```

Validació amb l'eina de Google Developers:

The screenshot shows the Google Developers Schema Validator interface. On the left, the HTML code is displayed with line numbers 1 through 16. Line 1 is highlighted with a black box, showing the opening tag for a 'Place' schema object: `<div itemscope itemtype="http://schema.org/Place">`. The right pane shows the validation results for the 'Place' object, including fields like `faxNumber`, `globalLocationNumber`, `hasMap`, `interactionCount`, `isicV4`, `logo`, `telephone`, and `alternateName`.

[Fig.13- Schema: depuració de l'objecte "Place".](#)

Anàlisi del codi:

Aquest objecte ajuda a representar un lloc amb totes les seves característiques (si es vol). En essència, com més microdades li adjuntem, més resultats obtindrem de l'objecte a cercar...

L'exemple, el qual em baso, és la meua ciutat Lleida i alguns dels seus trets més significatius, com; La Seu Vella (la Suda), El Centre Històric, l'Estació, etc.

L'ordre dels elements, que incloc dins del codi, és l'estructura que presenta Schema.org però, es poden posar de diverses maneres sempre que estiguin dins dels seus corresponents contenidors.

Utilitzant l'Schema.org, no es fa menester crear moltes subdivisions dels propis elements sinó que, quan insertem un element –tot seguit– li posem les seves pròpies característiques. D'aquesta manera, se li dona una definició exacta del propi element.

Els **elements de l'objecte "Place"**, són; **address** (per inserir l'adreça postal), **aggregateRating** (son les valoracions globals de la gent dins la xarxa), **containedIn** (especifica que un contenidor es troba dins d'un altre contenidor), **event** (per especificar quins esdeveniments es realitzen al lloc), **faxNumber** (inserir el número de fax), **geo** (Localització del lloc que es defineix per, coordenades i/o forma), **globalLocationNumber** (és el número global de localització), **hasMap** (per poder inserir un mapa dins la pròpia web o per url), **interactionCount** (son les

interaccions que han tingut els usuaris de la xarxa amb la nostra web), **isicV4**(és la Norma Internacional de Classificació Industrial de totes les Activitats Econòmiques), **logo** (és el logotip/imatge i/o url del lloc que s'està definint), **openingHoursSpecification** (son les especificacions de l'horari d'obertura), **photo** (és una imatge del lloc definit, aquest pot ser un objecte imatge o una fotografia), **review** (s'inserta si s'ha realitzat alguna revisió) i, per acabar, **telephone**(s'inserta un número de contacte, si fos menester).

Les **propietats de l' objecte "Thing"**, son; **additionalType** (inserir per incloure un determinat tipus de microdades, si escau), **alternateName** (inserir un nom alternatiu), **description** (poder realitzar una descripció de l'objecte), **image** (inserir una imatge, objecte imatge o url), **name** (inserir un nom identificatiu a l'objecte), **potentialAction** (per determinar una acció per realitzar, una acció potencial), **sameAs** (inserir una pàgina web que doni informació del lloc que estem definint), **url** (per inserir més d'una web que parli del lloc que estem definint).

- **"Person"(Persona):**

Aquest objecte és molt extens, amb molta diversitat tant d'elements com de propietats. És un dels objectes que s'utilitzen més ja que, per Internet, l' informació personal de cada un va sempre per xarxa...

L'exemple que exposo a continuació és una web definint-me, tota l' informació és del tot fictici (pel tema de protecció de dades).

CODI:

```
<div itemscope itemtype="http://schema.org/Person">
  <span itemprop="additionalName">Alias</span>
  <div itemprop="address" itemscope itemtype="http://schema.org/PostalAddress">
    <div itemprop="addressCountry" itemscope itemtype="http://schema.org/Country">
      <span itemprop="name">Catalunya</span>
    </div>
    <span itemprop="addressRegion">Lleida</span>
    <span itemprop="addressLocality">Lleida</span>
  </div>
  <div itemprop="affiliation" itemscope itemtype="http://schema.org/Organization">
    <span itemprop="name">Greenpeace</span>
    <span itemprop="url">http://www.greenpeace.org/espana/es/</span>
  </div>
  <div itemprop="alumniOf" itemscope
itemtype="http://schema.org/EducationalOrganization">
    <span itemprop="legalName">Universitat Oberta de Catalunya</span>
    <span itemprop="alternateName">UOC</span>
    <span itemprop="sameAs">www.uoc.edu</span>
  </div>
  <span itemprop="award">1r. Cursa Popular de Lleida 8500 mts.</span>
  <span itemprop="birthDate">1975-10-12</span>
  <div itemprop="birthPlace" itemscope itemtype="http://schema.org/Place">
    <div itemprop="address" itemscope itemtype="http://schema.org/PostalAddress">
      <div itemprop="addressCountry" itemscope itemtype="http://schema.org/Country">
```

```

    <span itemprop="name">Lleida</span>
  </div>
  <span itemprop="addressRegion">Lleida</span>
  <span itemprop="addressLocality">Lleida</span>
</div>
</div>
<div itemprop="brand" itemscope itemtype="http://schema.org/Brand">
  <span itemprop="name">AQUARIUS, Botella 1l.</span>
</div>
<div itemprop="brand" itemscope itemtype="http://schema.org/Organization">
  <span itemprop="name">Empresa Multinacional de progamacio en 3D</span>
  <span itemprop="alternateName">3D-LleidaGames</span>
  <div itemprop="address" itemscope itemtype="http://schema.org/PostalAddress">
    <div itemprop="addressCountry" itemscope itemtype="http://schema.org/Country">
      <span itemprop="name">Lleida</span>
    </div>
    <span itemprop="addressRegion">Mollerussa</span>
    <span itemprop="addressLocality">C/ Urgell num.77</span>
  </div>
</div>
</div>
<div itemprop="children" itemscope itemtype="http://schema.org/Person">
  <span itemprop="givenName">No tinc fills.</span>
</div>
<div itemprop="colleague" itemscope itemtype="http://schema.org/Person">
  <span itemprop="givenName">Nom amic</span>
</div>
<div itemprop="contactPoint" itemscope
itemtype="http://schema.org/ContactPoint">
  <span itemprop="contactType">technical support</span>
  <span itemprop="name">Google+</span>
  <span itemprop="url">https://plus.google.com/+DaniVillanueva/</span>
</div>
<span itemprop="url">https://plus.google.com/+DaniVillanueva/</span>
<span itemprop="deathDate">9999-12-31</span>
<div itemprop="deathPlace" itemprop itemtype="http://schema.org/Place">
  <span itemprop="name">GandalphElGris</span>
</div>
<span itemprop="duns">123456789</span>
<span itemprop="email">dani_est_uoc@gmail.com</span>
<span itemprop="familyName">Dani</span>
<span itemprop="faxNumber">123-456-7890</span>
<div itemprop="follows" itemscope itemtype="http://schema.org/Person">
  <span itemprop="givenName">Dani</span>
</div>
<span itemprop="gender">Hombre</span>
<span itemprop="givenName">Daniel Villanueva</span>
<span itemprop="globalLocationNumber">1234567890123</span>
<div itemprop="hasPOS" itemscope itemtype="http://schema.org/Place">
  <span itemprop="name">Es el lloc on treballa.</span>
</div>

```



```

<span itemprop="height">1.88 m</span>
<div itemprop="height" itemscope itemtype="http://schema.org/QuantitativeValue">
  <span itemprop="value">188</span>
  <span itemprop="unitCode">cm</span>
</div>
<div itemprop="homeLocation" itemscope itemtype="http://schema.org/Place">
  <div itemprop="address" itemscope itemtype="http://schema.org/PostalAddress">
    <div itemprop="addressCountry" itemscope itemtype="http://schema.org/Country">
      <span itemprop="name">Lleida</span>
    </div>
    <span itemprop="addressRegion">Lleida</span>
    <span itemprop="addressLocality">Lleida</span>
  </div>
</div>
<span itemprop="honorificPrefix">Eng.</span>
<span itemprop="honorificSuffix">Informatic</span>
<span itemprop="interactionCount">26 Seguidores, 50,321 vistas</span>
<span itemprop="isicV4">num. Norma Internacional</span>
<span itemprop="jobTitle">Eng. Tecnic. Informatic + Dissenyador 3D</span>
<div itemprop="knows" itemprop itemtype="http://schema.org/Person">
  <span itemprop="givenName">Steve Jobs</span>
</div>
<div itemprop="makesOffer" itemscope itemtype="http://schema.org/Offer">
  <span itemprop="name">Les meves fotos:</span>
  <span itemprop="url">https://picasaweb.google.com/dani</span>
</div>
<div itemprop="memberOf" itemscope
itemtype="http://schema.org/ProgramMembership">
  <span itemprop="name">Lleida</span>
</div>
<div itemprop="memberOf" itemscope
itemtype="http://schema.org/ProgramMembership">
  <span itemprop="name">3D-LleidaGames</span>
</div>
<span itemprop="naics">1234567890</span>
<div itemprop="nationality" itemscope itemtype="http://schema.org/Country">
  <span itemprop="name">Lleida</span>
</div>
<div itemprop="netWorth" itemscope
itemtype="http://schema.org/PriceSpecification"> <!-- TONTERÍA, ¿alguien tiene
precio?, pero nimodo, ellos lo ponen -->
  <span itemprop="Price">llimitat</span>
</div>
<div itemprop="owns" itemscope itemtype="http://schema.org/Product">
  <span itemprop="name">AQUARIUS 100% Natural</span>
</div>
<div itemprop="owns" itemscope itemtype="http://schema.org/OwnershipInfo">
  <span itemprop="name">AQUARIUS 100% Natural</span>
</div>
<div itemprop="parent" itemscope itemtype="http://schema.org/Person">

```

```

<span itemprop="givenName">Dani Villanueva</span>
<span itemprop="gender">Home</span>
</div>
<div itemprop="parent" itemscope itemtype="http://schema.org/Person">
  <span itemprop="givenName">Patricia Puigverd</span>
  <span itemprop="gender">Dona</span>
</div>
<div itemprop="performerIn" itemscope itemtype="http://schema.org/Event">
  <span itemprop="name">Las Reuniones en 3D-LleidaGames</span>
  <div itemprop="location" itemscope itemtype="http://schema.org/Place">
    <span itemprop="name">El 3D-LleidaGames</span>
    <div itemprop="address" itemscope itemtype="http://schema.org/PostalAddress">
      <span itemprop="name">Edificio del 3D-LleidaGames</span>
      <div itemprop="addressCountry" itemscope itemtype="http://schema.org/Country">
        <span itemprop="name">Lleida</span>
      </div>
      <span itemprop="addressRegion">Mollerussa</span>
      <span itemprop="addressLocality">C/ Urgell num.77</span>
    </div>
  </div>
  <span itemprop="startDate">2016-05-11</span>
  <div itemprop="offers" itemscope itemtype="http://schema.org/Offer">
    <span itemprop="name">Projectes a mesura</span>
    <span itemprop="price">2500€</span>
    <span itemprop="url">http://www.3D-LleidaGaames.html</span>
  </div>
</div>
<div itemprop="relatedTo" itemscope itemtype="http://schema.org/Person">
  <span itemprop="givenName">Pol Villanueva</span>
  <span itemprop="potentialAction">Vesabi</span>
</div>
<div itemprop="seeks" itemscope itemtype="http://schema.org/Demand">
  <span itemprop="url">https://picasaweb.google.com/dani</span>
</div>
<div itemprop="sibling" itemscope itemtype="http://schema.org/Person">
  <span itemprop="givenName">Pol Villanueva</span>
</div>
<div itemprop="spouse" itemscope itemtype="http://schema.org/Person">
  <span itemprop="givenName">No en tinc</span>
</div>
<span itemprop="taxID">DNI</span>
<span itemprop="telephone">123456789</span>
<span itemprop="vatID">21% IVA</span>
<div itemprop="weight" itemscope itemtype="http://schema.org/QuantitativeValue">
  <span itemprop="value">80</span>
  <span itemprop="unitCode">Kg</span>
</div>
<div itemprop="workLocation" itemscope itemtype="http://schema.org/Place">
  <span itemprop="name">Casa meva</span>
  <div itemprop="address" itemscope itemtype="http://schema.org/PostalAddress">

```

```

<span itemprop="name">Casa</span>
<div itemprop="addressCountry" itemscope itemtype="http://schema.org/Country">
  <span itemprop="name">Lleida</span>
</div>
<span itemprop="addressRegion">Mollerussa</span>
<span itemprop="addressLocality">C/ Urgell num.77</span>
</div>
</div>
<div itemprop="worksFor" itemscope itemtype="http://schema.org/Organization">
  <span itemprop="name">GreenPeace, Dreamorks</span>
  <span itemprop="description">Realitzant tasques</span>
</div>
<span itemprop="alternateName">pinpin</span>
<span itemprop="description">Persona que es diu, Alex</span>
<span itemprop="image">http://www.abc.es/Media/201503/30/hombre-mas-guapo--478x350.jpg</span>
<div itemprop="image" itemscope itemtype="http://schema.org/ImageObject">
  <span itemprop="name">hombre-mas-guapo--478x350.jpg</span>
  <span itemprop="url">http://www.abc.es/Media/201503/30/hombre-mas-guapo--478x350.jpg</span>
</div>
<span itemprop="name">Dani</span>
<span itemprop="potentialAction">Realitzar el TFM</span>
<span itemprop="sameAs">www.danivillanueva.person</span>
<span itemprop="url">http://plus.google.com/+DaniVillanueva/</span>
</div>

```

Validació amb l'eina de Google Developers:

The screenshot shows the Google Developers Schema Validator interface. On the left, the JSON-LD code is displayed with line numbers 1 through 15. The code defines a 'Person' schema with various properties like 'additionalName', 'address', 'affiliation', and 'url'. On the right, the 'Resultados' section shows the extracted data for the 'Person' type, including 3 warnings. The data is presented in a table format with the following fields:

Person	
additionalName:	Alias
award:	1r. Cursa Popular de Lleida 8500 mts.
birthDate:	1975-10-12
url:	https://plus.google.com/+DaniVillanueva/
deathDate:	9999-12-31
name:	GandalphElGris
duns:	123456789
email:	dani_est_uoc@gmail.com
familyName:	Dani
faxNumber:	123-456-7890
gender:	Hombre
givenName:	Daniel Villanueva
globalLocationNumber:	1234567890123
honorificPrefix:	Eng.

Fig.14- Schema: depuració de l'objecte "Person".

Anàlisi del codi:

Aquest objecte ajuda a representar a una persona qualsevol amb totes les seves característiques (si es vol). En essència, com més microdades li adjuntem, més resultats obtindrem de la persona a cercar...

L'exemple, el qual em baso, és la meva pròpia persona amb informació fictícia.

L'ordre dels elements, que incloc dins del codi, s'estructura com ho presenta Schema.org però, es poden posar de diverses maneres sempre que estiguin dins dels seus corresponents contenidors.

Això, ens permetrà que ens trobin amb molta més facilitat, i molt més ràpid, quan facin una cerca per Internet. La raó serà perquè en cap altre lloc es parlarà tant extensament de la persona que s'esmenta dins la web. Com més s'estengui l'objecte "Persona" dins la pàgina o d'altres pàgines, més informació donarà als cercadors...

Els elements de l'objecte "Person", són; **additionalName**(per inserir el segon nom), **address**(és l'adreça postal), **affiliation**(si es forma part d'alguna afiliació), **alumniOf**(s'especifica d'on va ser alumne la persona que s'està definint, per poder utilitzar aquest element, primer s'ha d'inserir l'organització estudiantil i després, la persona), **award**(inserir els premis obtinguts), **birthDate**(data de naixement), **birthPlace**(lloc de naixement), **brand**(si la persona és propietària d'una organització i/o marca), **children**(si la persona té fills), **colleague**(és el colega de la persona que es defineix), **contactPoint**(és un punt de contacte), **deathDate**(si la persona està morta, es posa la seva data de defunció), **deathPlace**(lloc de la mort, lloc del repòs), **duns**(si es té, és el número d'identificació DUNS), **email**(correu electrònic), **familyName**(Cognom "Patern" o "Matern", si s'utilitza amb "givenName" es converteix automàticament en "Cognom Matern"), **faxNumber**(número de fax), **follows**(quina relació social té amb la persona que estem definint; Amic, Novia, etc.), **gender**(sexe de la persona), **givenName**(seria l'aliès de la persona però, depèn del país en que s'estigui), **globalLocationNumber**(número global d'identificació), **hasPOS**(és el punt de contacte, en Ventas), **height**(alçada de la persona), **homeLocation**(ubicació on es troba la casa, la llar), **honorificPrefix**(si es menester el prefix honorífic de la titulació, per exemple "doctor"), **honorificSuffix**(seria el segon esment honorífic de la titulació, si escau), **interactionCount**(número total d'interaccions dels usuaris a l'enllaç de la persona), **isicV4**(Núm. Norma Int. De Classificació Industrial de totes les Activitats Econòmiques), **jobTitle**(títol del seu treball, la seva professió o lloc de treball), **knows**(és la relació de treball que té; cap, gerent, subordinat, treballador, etc.), **makesOffer**(és un apuntador als productes que desenvolupa i/o serveis que ofereix), **memberOf**(és membre d'una Organització), **naics**(no faria falta posar-lo ja que només s'utilitza a Nord-Amèrica, és el número del Sistema de Classificació Industrial Americà), **nationality**(Nacionalitat), **netWorth**(és el valor econòmic que se li dona a l'objecte, en aquest cas a una persona, en la meua opinió no li tindria que ser-hi), **owns**(són els productes de la persona, o propietats que tingui), **parent**(són els pares de la persona), **performerIn**(si actua, dins d'un advert, com a artista o participant), **relatedTo**(relació familiar; padrí, tiet, cosí, etc.), **seeks**(apuntador cap a les demandes de productes o serveis sol·licitats per la persona), **sibling**(Germà), **spouse**(marit), **taxID**(Número d'Identificació Fiscal), **telephone**(telèfon), **vatID**(Identificació fiscal de l'IVA), **weight**(és el pes de la persona),

workLocation(és el lloc de treball de la persona), **worksFor**(son les organitzacions en les que treballa la persona que es defineix).

Les propietats de l' objecte "Thing", son; **additionalType** (inserir per incloure un determinat tipus de microdades, si escau), **alternateName** (inserir un nom alternatiu), **description** (poder realitzar una descripció de la persona), **image** (inserir una imatge de la persona o url), **name**(inserir un nom identificatiu de la persona), **potentialAction** (és l'acció potencial que realitza la persona), **sameAs**(inserir una pàgina web que doni informació de la persona que estem definint), **url**(per inserir més d'una web que parli dels elements que estem definint).

- "Organization" (organització):

Amb aquest objecte passa el mateix que li passava a l'anterior, "Person". Conté tantes microdades per poder definir qualsevol organització que existeixi. Per tant, és lògic que tingui diversos elements per poder classificar específicament qualsevol d'elles.

L'exemple, per poder analitzar aquest objecte, es basa en un restaurant fictici de Lleida...

CODI:

```
<div itemscope itemtype="http://schema.org/Organization">
  <span itemprop="name">Restaurant Marisc de Lleida</span>
  <span itemprop="alternateName">el marisc de Lleida</span>
  <span itemprop="description">Restaurant especialitzat en mariscades</span>
  <span itemprop="image">http://www.mariscdeLleida.com?elmarisc.jpg</span>
  <div itemprop="image" itemscope itemtype="http://schema.org/ImageObject">
    <span itemprop="name">http://www.mariscdeLleida.com?elmarisc.jpg</span>
  </div>
  <span itemprop="potentialAction">Menjar</span>
  <span itemprop="url">http://www.mariscdeLleida.com/qui_som.html</span>
  <span itemprop="sameAs">www.mariscdeLleida.com</span>
  <span itemprop="address">Lleida, Catalunya</span>
  <div itemprop="aggregateRating" itemscope
itemtype="http://schema.org/AggregateRating">
  <span itemprop="ratingValue">4.5 estrelles de 789 vots</span>
  <span itemprop="ratingCount">789</span>
</div>
  <div itemprop="brand" itemscope itemtype="http://schema.org/Brand">
  <span itemprop="name">Mariscada de tres plantes</span>
</div>
  <div itemprop="brand" itemscope itemtype="http://schema.org/Organization">
  <span itemprop="name">Marisqueries de Catalunya S.L.</span>
</div>
  <div itemprop="contactPoint" itemscope itemtype =
"http://schema.org/ContactPoint">
  <span itemprop="name">Lleida, Catalunya</span>
  <span itemprop="contactType">reservations</span>
</div>
```

```

<span itemprop="telephone">+43-(973)123456</span>
<span itemprop="url">http://www.mariscdeLleida.com/reserves.html</span>
</div>
<div itemprop="department" itemscope itemtype="http://schema.org/Organization">
  <span itemprop="name">Ventres</span>
</div>
<span itemprop="dissolutionDate">2016-05-11T12:00:00</span>
<span itemprop="duns">00000000000000</span>
<span itemprop="email">reserves@mariscdeLleida.com</span>
<div itemprop="employee" itemscope itemtype="http://schema.org/Person">
  <span itemprop="name">Daniel Villanueva</span>
</div>
<div itemprop="event" itemscope itemtype="http://schema.org/Event">
  <span itemprop="name">Concurs anual de pesca al riu Segre</span>
  <div itemprop="location" itemscope itemtype="http://schema.org/Place">
    <span itemprop="name">Lleida, Catalunya</span>
    <span itemprop="address">C/ Ferran num.10 (Baixos)</span>
  </div>
  <span itemprop="startDate">2016-05-11T12:30:00</span>
  <div itemprop="offers" itemscope itemtype="http://schema.org/Offer">
    <span itemprop="name">Mariscades de tres plantes al 50%</span>
    <span itemprop="price">24€</span>
    <span itemprop="url">http://www.mariscdeLleida.com/ofertes.html</span>
  </div>
</div>
<span itemprop="faxNumber">123-456-7890</span>
<div itemprop="founder" itemscope itemtype="http://schema.org/Person">
  <span itemprop="name">Dani Villanueva</span>
</div>
<div itemprop="foundingLocation" itemscope itemtype="http://schema.org/Place">
  <span itemprop="name">Lleida, Catalunya</span>
</div>
<span itemprop="globalLocationNumber">1234567890123</span>
<div itemprop="hasPOS" itemscope itemtype="http://schema.org/Place">
  <span itemprop="name">Restaurant Marisc de Lleida</span>
</div>
<span itemprop="interactionCount">115 likes - 21 shares</span>
<span itemprop="isicV4">4685</span>
<span itemprop="legalName">Restaurant Marisc de Lleida S.A.</span>
<div itemprop="location" itemscope itemtype="http://schema.org/Place">
  <span itemprop="name">Davant estació "dels Pirineus"</span>
</div>
<div itemprop="logo" itemscope itemtype="http://schema.org/ImageObject">
  <span itemprop="name">logo.jpg</span>
</div>
<div itemprop="makesOffer" itemscope itemtype="http://schema.org/Offer">
  <span itemprop="name">Plata d'assortiments de mariscs</span>
  <span itemprop="price">79€</span>
  <span itemprop="url">http://www.mariscdeLleida.com/ofertes.html</span>
</div>

```

```

<div itemprop="member" itemscope itemtype="http://schema.org/Person">
  <span itemprop="name">Marta Vicens</span>
</div>
<div itemprop="member" itemscope itemtype="http://schema.org/Organization">
  <span itemprop="name">Seafood School of California</span>
</div>
<div itemprop="memberOf" itemscope itemtype="http://schema.org/Organization">
  <span itemprop="name">International Seafood org.</span>
</div>
<div itemprop="memberOf" itemscope
itemtype="http://schema.org/ProgramMembership">
  <span itemprop="name">Els mariscos catalans</span>
</div>
  <span itemprop="naics">000000000</span>
<div itemprop="owns" itemscope itemtype="http://schema.org/Product">
  <span itemprop="name">Mariscades de tres plantes</span>
</div>
<div itemprop="owns" itemscope itemtype="http://schema.org/OwnershipInfo">
  <span itemprop="name">Restaurant Marisc de Lleida</span>
</div>
<div itemprop="review" itemscope itemtype="http://schema.org/Review">
  <span itemprop="name">Neteja ref: #798SF</span>
  <div itemprop="author" itemscope itemtype="http://schema.org/Person">
    <span itemprop="name">Daniel Villanueva</span>
  </div>
</div>
<div itemprop="seeks" itemscope itemtype="http://schema.org/Demand">
  <span itemprop="name">Demanda 8074</span>
</div>
<div itemprop="subOrganization" itemscope
itemtype="http://schema.org/Organization">
  <span itemprop="name">L'ecosistema submari</span>
</div>
  <span itemprop="taxID">73.789.456G</span>
  <span itemprop="telephone">+43-(973)123456</span>
  <span itemprop="vatID">73.789.456G</span>
</div>

```

Validació amb l'eina de Google Developers:

The screenshot shows the Google Developers Schema Validator interface. On the left, there is a code editor with the following HTML code:

```
1 <div itemscope itemtype="http://schema.org/Organization">
2 <span itemprop="name">Restaurant Marisc de Lleida</span>
3 <span itemprop="alternateName">el marisc de Lleida</span>
4 <span itemprop="description">Restaurant especialitzat en
mariscades</span>
5 <span itemprop="image">http://www.mariscdeLleida.com?
elmarisc.jpg</span>
6 <div itemprop="image" itemscope
itemtype="http://schema.org/ImageObject">
7 <span itemprop="name">http://www.mariscdeLleida.com?
elmarisc.jpg</span>
8 </div>
9 <span itemprop="potentialAction">Menjar</span>
10 <span
itemprop="url">http://www.mariscdeLleida.com/qui_som.html</span>
11 <span itemprop="sameAs">www.mariscdeLleida.com</span>
12 <span itemprop="address">Lleida, Catalunya</span>
13 <div itemprop="aggregateRating" itemscope
itemtype="http://schema.org/AggregateRating">
```

On the right, the results pane shows the extracted JSON-LD data for the Organization:

Organization (1)	
name:	Restaurant Marisc de Lleida
alternateName:	el marisc de Lleida
description:	Restaurant especialitzat en mariscades
image:	http://www.mariscdeLleida.com?elmarisc.jpg
url:	http://www.mariscdeLleida.com/qui_som.html
sameAs:	www.mariscdeLleida.com
dissolutionDate:	2016-05-11T12:00:00
duns:	0000000000000
email:	reserves@mariscdeLleida.com
faxNumber:	123-456-7890
globalLocationNumber:	1234567890123
interactionCount:	115 likes - 21 shares
isicV4:	4685
legalName:	Restaurant Marisc de Lleida S.A.

Fig.15- Schema: depuració de l'objecte "Organization".

Anàlisi del codi:

Aquest objecte ajuda a representar a una determinada organització (en el nostre cas, un restaurant dedicat als mariscs). Aquesta entitat s'ha de definir completament, amb totes les seves característiques (si es vol). En essència, com més microdades li adjuntem, més resultats obtindrem del restaurant a cercar...

L'exemple, el qual em baso, és un restaurant qualsevol de mariscs de Lleida, amb informació fictícia. Per poder definir l'objecte, podia optar per dos vessants:

La primera, des del tipus:

"Thing/Organization/LocalBusiness/FoodEstablishment/Restaurant".

I la segona des del tipus:

"Thing/Place/LocalBusiness/FoodEstablishment/Restaurant".

Com es pot apreciar, la ruta és quasi be la mateixa, l'únic que canvia és l'objecte principal a que farà referència aquella pàgina web. Depenent del significat que hagi d'agafar aquesta, optarem per una o per una altra. Si hagués de tenir dels dos objectes, hauríem d'incloure els vocabularis d'aquestes a la pàgina a definir.

Aquesta part tant "versàtil" origina a que es puguin realitzar infinitats de possibles combinacions amb els respectius objectes, fent que cada definició d'objecte sigui molt acurada i única. Al haver agafat el primer camí per l'objecte restaurant (via "Organization"), em dona l'opció de poder "direccionar" el restaurant cap a possibles organitzacions, les quals pugui pertànyer...

L'ordre dels elements, que incloc dins del codi, s'estructura com ho presenta Schema.org però, es poden posar de diverses maneres sempre que estiguin dins dels seus corresponents contenidors sinó, l'eina de depuració del Google no reconeix la sentència dins del contenidor erroni.

Això, ens permetrà que ens trobin amb molta més facilitat, i molt més ràpid, quan facin una cerca per Internet. La raó serà perquè en cap altre lloc es parlarà tant

extensament del restaurant que s'esmenta dins la web. Com més s'estengui l'objecte "Organització" dins la pàgina o d'altres pàgines, més informació donarà als cercadors del restaurant creat dins d'aquesta organització...

Els elements de l'objecte "Organization", son; **address**, **aggregateRating**, **brand**(és la marca associada a un producte o un servei de la pròpia organització), **contactPoint** (és un punt de contacte), **department**(és un departament de la pròpia organització), **dissolutionDate**(si existeix un dia de dissolució de l'organització), **duns**(Identificació DUNS, si escau), **email**(correu electrònic), **employee**(treballador), **events**(si escau algun advent de l'organització), **faxNumber**, **founder**(si hi ha un fundador), **foundingDate**(la data en que es va fundar l'organització), **foundingLocation**(lloc on es va fundar l'Organització), **globalLocationNumber**, **hasPOS**(és el punt de vendes gestionada per l'organització), **interactionCount**, **isicV4**, **legalName**(nom real/legal de l'Organització), **location**(ubicació, lloc o direcció postal), **logo**, **makesOffer**(Apuntador cap a les ofertes, productes i/o serveis que ofereix l'Organització), **member**(són membre/s de l'Organització o la persona en sí), **memberOf**(és membre de una organització o d' un altre grup a destacar), **naics**(sistema de classificació Industrial de Nord-America), **Owns**(l'Organització té un producte o un valor estructurat del producte), **review**, **seeks**(si hi ha cap demanda), **subOrganization**(si hi ha cap organització dins d'una altra organització), **taxID**(Número Identificació Fiscal), **telephone**(número de telèfon), **vatID**(identificador de l' IVA), **additionalType**, **alternateName**, **description**, **image**, **name**, **potentialAction**, **sameAs** i **url**.

The Open Graph Protocol:

The Open Graph Protocol (OGP), permet a qualsevol pàgina web, convertir-se en un objecte ric dins d'una xarxa/àmbit social. L' OGP s'utilitza moltíssim dins de les xarxes socials, sobretot Facebook. Aquesta, l'utilitza per poder tenir qualsevol objecte dins de la pròpia web , la mateixa funcionalitat de qualsevol objecte dins de la pròpia xarxa social.

Gràcies a Facebook, l' OGP ha tingut molta més acceptació dins de les xarxes socials, ja que permet definir com es vol que es publiqui el contingut del compte web, dins de Facebook.

Per a que l'OGP funcioni en una web, s' han d' introduir una sèrie d'etiquetes en el codi d' aquesta. Començaríem per les etiquetes de capçalera(<head></head>),les quals dependrien de; la tipologia d'informació a publicar, descripcions, imatges i els enllaços(url). Tanmateix, Open Graph Protocol, funciona amb etiquetes html(<meta>) i amb la definició de la propietat "og:" .

```
<meta property="og:title" content="Títol de l'article dins Facebook.">
```

[Fig.16- OGP: Etiquetatge HTML, propietat "og:title".](#)

Si aquesta etiqueta no hi fos, Facebook agafaria automàticament el títol del contingut a publicar.


```
<meta property="og:description" content="Descripció de la web per  
publicar dins Facebook.">
```

[Fig.17- OGP: Etiquetatge HTML, propietat "og:description".](#)

Si aquesta etiqueta no hi fos, Facebook agafaria la descripció destinada al "SEO" i, per falta d'aquesta, agafa el primer paràgraf del post o del propi contingut. S'ha de tenir en compte que, com a màxim, "og:description", només pot emmagatzemar 195 caràcters.

```
<meta property="og:img" content="http://www.lamevaweb.es/img/foto_perfil.jpg">  
<!-- url Ficticia -->
```

[Fig.18- OGP: Etiquetatge HTML, propietat "og:img".](#)

La imatge, ha de tenir una mida no superior als 200px*200px. Si l'etiqueta no existís, Facebook dona l'opció que l'usuari trii d'entre les que es trobin dins de la seva web. Per si fos menester, i s'hagués de posar més d'una imatge, es dona l'opció de copiar –reiteradament- l'última sentència esmentada (només cal canviar l'url de la imatge a tractar).

Arrivats en aquest punt, necessitaríem etiquetes complementàries d'informació al propi compte de Facebook, però sense ser visibles per a l'usuari.

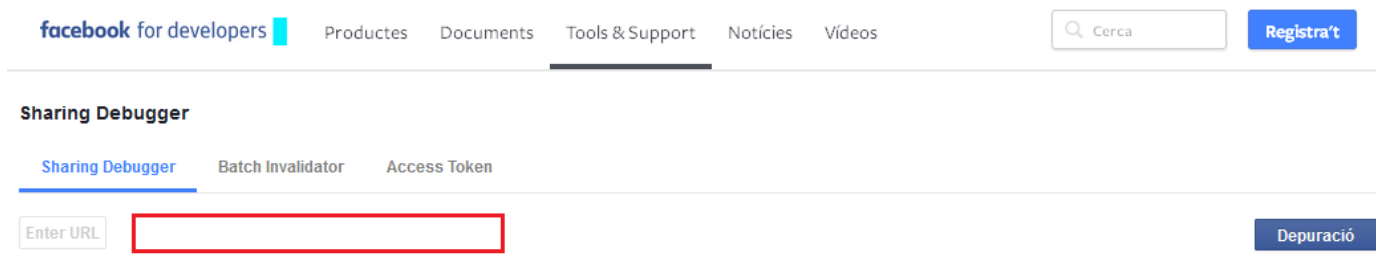
Aquestes, anirien precedides per "og:site_name" o "og:locale". La primera, comunica la ubicació original de la informació. La segona, especifica l'idioma amb el que es troba escrit l'informació. Per tant, un bon exemple, seria:

```
<meta property="og:site_name" content="http://www.danivillanueva.com">  
<!-- url Ficticia -->  
<meta property="og:locale" content="es_ES">
```

[Fig.19- OGP: Etiquetatge HTML, propietat "og:site name" i "og:locale".](#)

Igual que Schema.org, Facebook ofereix una manera de poder depurar les etiquetes OGP i, d'aquesta manera poder avaluar si funcionen correctament.

L'enllaç per accedir-hi és, <https://developers.facebook.com/tools/debug>.



[Fig.20- OGP: Depurador d'etiquetes OGP, per Facebook.](#)

Per implementar l'explicació donada més a dalt, ens centrarem en una pàgina "real", un restaurant de Lleida:



[Fig.21- Pàgina web del "Restaurant Nou Casa José".](#)

Si mirem la seva estructura interna, ens trobarem:

```
<!DOCTYPE html>
<html lang="es">
<head>
<meta http-equiv="Content-Type" content="text/html; charset=UTF-8" />
<meta name="viewport" content="width=device-width, initial-scale=1, maximum-
scale=1" />
<meta property="og:title" content="Restaurante en Lleida con Restaurant Nou
Casa
José" />
<meta name="description" content="En nuestro prestigioso restaurante degustará
las
más deliciosas especialidades gastronómicas.
Estamos situados en la localidad de Lleida. No
dude en visitarnos." />
<meta property="og:description" content="En nuestro prestigioso restaurante
degustará las más deliciosas
especialidades gastronómicas. Estamos
situados en la localidad de Lleida. No
dude en visitarnos." />
<meta name="generator" content="Websplanet" />
<meta name="robots" content="index, follow" />
<title>Restaurante en Lleida con Restaurant Nou Casa José</title>
```




```
<link  
href="http://static.miweb.paginasamarillas.es/templates/resellers/1001/favicon.ico?v=  
" rel="SHORTCUT ICON" type="image/x-icon" />
```

...

```
<meta name="enterprise" content=" RESTAURANT NOU CASA JOSÉ">  
</head>  
<body>  
.....  
</body>  
</html>
```

Com es pot veure, tot passa a la part de la capçalera de la web (“<header>”). Tota la configuració de l’OGP pren part abans que es carregui el cos de la web. Abans de tenir la informació, ha de saber a ón col·locar-la...

Si aquesta web la insertem al “Depurador del Facebook”, aquest ens demanarà – solament- l’ url de la web. Al fer-ho, ens depura la web i com a resultat, ens surt el següent:

When and how we last scraped the URL	
Time Scraped	fa 9 segons Scrape Again
Response Code	206
Fetches URL	http://www.restaurantcasajose.com/
URL canònica	http://www.restaurantcasajose.com/ 16 likes, shares and comments (Més informació)
Link Preview	 <p>Restaurante en Lleida con Restaurant Nou Casa José En nuestro prestigioso restaurante degustará las más deliciosas especialidades gastronómicas. Estamos situados en la localidad de Lleida. No dude en visitarnos. RESTAURANTCASAJOSE.COM</p>
Server IP	54.77.100.126

[Fig.22- Depuració de la web “Casa José”, per l'eina “Facebook for Developers”\(1\).](#)

Based on the raw tags, we constructed the following Open Graph properties

fb:app_id	
og:url	http://www.restaurantcasajose.com/
og:type	website
og:title	Restaurante en Lleida con Restaurant Nou Casa José
og:description	En nuestro prestigioso restaurante degustará las más deliciosas especialidades gastronómicas. Estamos Lleida. No dude en visitarnos.
og:image	http://static.miweb.paginasamarillas.es/var/m_1/1d/1d2/13716/259237-banner2.12.jpg

```

<meta name="viewport" content="width=device-width, initial-scale=1, maximum-scale=1" />
<meta property="og:title" content="Restaurante en Lleida con Restaurant Nou Casa José" />
<meta name="description" content="En nuestro prestigioso restaurante degustará las más delicic
<meta property="og:description" content="En nuestro prestigioso restaurante degustará las más
<meta name="generator" content="Websplanet" />
<meta name="robots" content="index,follow" />
<meta name="enterprise" content=" RESTAURANT NOU CASA JOSÉ" />

```

URLs

Open Graph	Open Graph Object Debugger
Graph API	https://graph.facebook.com/10150771727301298
Scraped URL	See exactly what our scraper sees for your URL
Scrape API	Scrape via API

[Fig.23- Depuració de la web “Casa José”, per l’eina “Facebook for Developers”\(2\).](#)

Per tant, podem veure que, en OGP no tenim les estructures jeràrquiques tant extenses com les de l’Schema.org, per l’exemple del “Restaurant”. L’OGP, només necessita al voltant d’unes 10 metadates per poder estructurar la seva informació, però no queda del tot ben definida. En canvi, en Schema.org s’estructura molt bé l’informació i, a més, facilita la creació d’altres contenidors per “ampliar” la seva conceptualització. També val a dir que, necessita de moltes més etiquetes i codi que l’OGP, però és més estructurat i intuïtiu l’Schema.org que l’OGP. Aquest últim, disposa de poquíssimes etiquetes (en comparació amb Schema.org) i, a l’hora d’estructurar i definir, no ho fa tant a consciència com l’Schema.org. Sobretot, hem de tenir en compte que l’OGP permet a qualsevol web, convertir-se en un objecte dins d’un contingut social, és a dir, les xarxes socials. Facebook ha acceptat l’OGP per millorar la visibilitat i les publicacions externes del facebook. Per a que aquest etiquetatge funcioni dins del Facebook, s’han d’introduir un seguit d’etiquetes; les capçaleres(head), el tipus del contingut d’informació i les pertinents descripcions, imatges i enllaços(url).

Pertant, OGP té la següent estructura d'etiquetatge:

<meta property="og:..." content="">

Com a og:type, tenim:

og:title: <meta property="og:title" content="Títol de la publicació">

og:description:

<meta property="og:description" content="Descripció del lloc a publicar">

og:image: <meta property="og:image" content=<http://url de l'imatge>>

og:url: <meta property="og:url" content="http://domini.com/direccio-url-canonica">

Com a complementàries, tindríem:

og:site_name : Diu l'origen de l'informació al Facebook.

<meta property="og:site_name" content=<http://skynet.com>>

og:locale: Informa al Facebook sobre l'idioma en que es troba escrit el contingut de la publicació.

<meta property="og:locale" content="es_ES">

Per aquesta raó, hem de tenir en compte l'utilització de l'OGP. En el meu punt de vista, per publicacions de Facebook o twitter (i alguna altra xarxa social que se m'escapa), va molt bé perquè no és necessari gran volum de contextualització de les dades. Però, per pàgines webs i per l'emmagatzematge d'informació, prefereixo molt més l' Schema.org. És més complex –si mirem en el gran nombre d'objectes, elements i propietats que disposa- però, és més estructurat que l'OGP i més intuïtiu.

2.4 Anàlisi de les Microdades, Microformats i RDFas:

a. Estructura.

b. Comportament.

Es troben objectius comuns entre els RDFas i els Microformats. Mentre, en els Microformats especifiquen la sintaxis per poder introduir dades estructurades (en Html) per poder utilitzar –lo com un vocabulari propi, de termes específics per a cada microformat creat. En canvi, els RDFas només especifiquen la sintaxis necessària per a que, els diferents termes de diversos vocabularis, es puguin barrejar lliurement i que aquest, pugui ser analitzat sense el vocabulari original creat.

Microdades:

Les microdades van ser creades entre Google, Microsoft i Yahoo. Sorgeixen amb l'aparició d'HTML (versió 5). Aquests, poden ajudar a que el cercador ,que busca

una determinada informació, entengui i mostri –de manera estructurada- la informació que té la web. Aquesta manera d'etiquetatge abasta un ventall amplíssim de possibilitats de diverses tipologies de dades, com: les dades personals d'una persona, events musicals (concerts), geolocalització d'empreses (restaurants, hospitals, etc), etc...

Per poder estructurar tota aquesta diversitat d'informació, és necessari que el seu funcionament es basi en l'etiquetatge de l'informació per **nom-valor**. Per poder-ho realitzar, utilitza l'atribut **itemscope** per poder marcar l'informació què, més tard, serà etiquetada per l'atribut, **itemprop**.

La gran avantatja de poder treballar amb les microdades és poder convertir el text en informació entenedora per al Google o qualsevol altre cercador. Ja que, definim el significat de la informació, l'estructurem i la relacionem entre si. La seva gran simplicitat en la seva implementació fa que sigui més senzill i entenedor utilitzar les microdades que el RDFa, ja que és més complex. També, al ser compatible amb Schema.org, li dona una certa normalització en el seu etiquetatge i descripció dels propis ítems.

L'ús de les microdades no és gens intrusiu per al codi, ja que tot l'etiquetatge ve donat per Schema.org. La seva senzilla implementació es deguda a l' utilització d'atributs dins del codi HTML.

```
<div itemscope itemtype="http://schema.org/Organization">
  <h2 itemprop="name">UOC</h2>
  <strong>Email:</strong> <a href="mailto:info@uoc.edu" itemprop="email">info@uoc.edu</a>
  <br>
  <strong>Direccio:</strong> <span itemprop="telephone">Rambla Ferran, num. 24</span>
  (<span itemprop="addressLocality">Barcelona</span>)
  <br>
  <strong>Telefon:</strong> <span itemprop="telephone">(93) 7324564</span>
  <div itemprop="geo" itemscope itemtype="http://schema.org/GeoCoordinates">
    <meta itemprop="latitude" content="40.20" />
    <meta itemprop="longitude" content="-4.50" />
  </div>
</div>
```

[Fig.24: Exemple senzill d'una codificació i estructuració d'una pàgina web en HTML i amb microdades.](#)

Al executar el codi, mitjançant el navegador **iExplorer**, obtenim la següent pàgina :

UOC

Email: info@uoc.edu

Direccio: Rambla Ferran, num. 24 (Barcelona)

Telefon: [\(93\)7324564](tel:937324564)

[Fig.25: Execució del codi HTML amb microdades.](#)

Veient el resultat podem dir que, la utilització de les microdades és molt senzill, intuïtiu i res d'intrusiu. A més, els podem utilitzar per multitud d'estructures, aquestes compatibles amb **Schema.org**. Per finalitzar, podem corregir l'estructura i la semàntica de la nostra web tot anant a l'eina, que ens ofereix Google, de **verificació de les dades estructurades**.

Avui en dia, les microdades d'Schema.org ja comencen a mostrar snippets enriquits en Google. La seva bona utilització i acceptació a *Android Market*, fa que ens pensem que Google anirà interpretant més estructures d'Schema.org fins que les tingui totes... Això no influirà negativament als altres tipus de marcatges però si – indirectament- amb el ranking de posicionament de les pàgines web pels seus continguts i etiquetatge. Penso que Google diu que el fet de tenir microdades amb Schema dins la web, no “influeix” al ranking dels resultats de les pròpies cerques... No obstant, és lògic pensar que d'alguna manera, les webs amb microdades es puguin “beneficiar” de pujar dins del ranking, ja que en certa manera, l'estructuració de l'informació de la pàgina pot facilitar l'interpretació dels continguts d'aquesta i, per tant, els cercadors trobin més ràpidament l'informació que busquen, afavorint – indirectament- el posicionament de la web dins del ranking.

Microformats:

Els microformats estan formats per codi HTML o XHTML. A més, estan compostos d'una sèrie d'atributs o classes CSS, aquests per poder identificar l'informació que contenen. Tanmateix, proporcionen significat semàntic al contingut per a que, pugui ser entès i comprès per les màquines que cerquin la respectiva informació.

Per poder programar amb ells, només s'ha d'insertar l'informació dins de classes, i aquestes, dins d'etiquetes HTML. Com per exemple, el següent codi ens mostra unes valoracions i/o comentaris, encapçalats per la classe “**hreview**”:

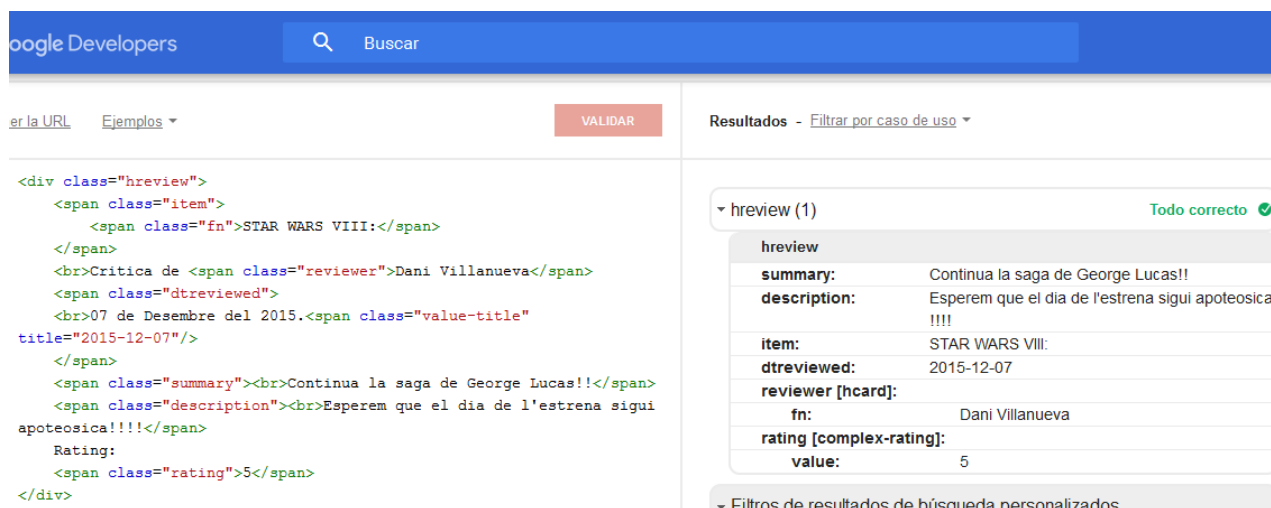
```
<div class="hreview">
  <span class="item">
    <span class="fn">STAR WARS VIII:</span>
  </span>
  <br>Critica de <span class="reviewer">Dani Villanueva</span>
  <span class="dtreviewed">
    <br>07 de Desembre del 2015.<span class="value-title" title="2015-12-07"/>
  </span>
  <span class="summary"><br>Continua la saga de George Lucas!!</span>
  <span class="description"><br>Esperem que el dia de l'estrena sigui apoteosica!!!!</span>
  Rating:
  <span class="rating">5</span>
</div>
```

[Fig.26: Exemple senzill d'una codificació i estructuració d'una pàgina web en HTML i amb microformats.](#)

Com es pot veure, cada contingut d'informació està encapsulada dins d'una etiqueta HTML amb una classe específica. La qual, determina quin contingut conté la pròpia

informació. Totes aquestes classes, s'engloben dins d'una classe principal, anomenada **“hreview”** quina atorga la condició de microformat al contingut que inclou.

Si el codi, introduït anteriorment, el validem amb l'eina **“Structured Data Testing Tool”**, obtenim la següent validació:



```
<div class="hreview">
  <span class="item">
    <span class="fn">STAR WARS VIII:</span>
  </span>
  <br>Critica de <span class="reviewer">Dani Villanueva</span>
  <span class="dtreviewed">
    <br>07 de Diciembre del 2015.<span class="value-title"
title="2015-12-07"/>
  </span>
  <span class="summary"><br>Continua la saga de George Lucas!!</span>
  <span class="description"><br>Esperem que el dia de l'estrena sigui
apoteosica!!!!</span>
  Rating:
  <span class="rating">5</span>
</div>
```

hreview	
summary:	Continua la saga de George Lucas!!
description:	Esperem que el dia de l'estrena sigui apoteosica!!!!
item:	STAR WARS VIII:
dtreviewed:	2015-12-07
reviewer [hcard]:	
fn:	Dani Villanueva
rating [complex-rating]:	
value:	5

Fig.27: Testeig d'una pàgina web en HTML amb l'eina “Structured Data Testing Tool”, de Google.

Al executar el codi, mitjançant el navegador **Firefox**, obtenim la següent pàgina:

STAR WARS VIII:
Critica de Dani Villanueva
07 de Diciembre del 2015.
Continua la saga de George Lucas!!
Esperem que el dia de l'estrena sigui apoteosica!!!! Rating: 5

Fig.28: Execució del codi HTML amb microformats.

Amb aquest programa, es pot testejar la correcció dels nostres Microformats així com l'aspecte que aquests tindran en els **SERPs**(Pagina de Resultats de Cerca, **“Search Engine Result Page”**).

Una altra particularitat, seria la combinació de microformats. Utilitzen conjunts d'estructures simples de les propietats de cada microformat, a on dona la possibilitat de que múltiples d'aquests es combinin per poder crear /especificar una estructura adicional. Com a exemple, podríem anomenar que “Events”, sol tenir sempre l'informació local amb una estructura adicional, com la Direcció.

Una de les avantatges d'utilitzar-los es que s'utilitzen per els motors de cerca i poden optimitzar les pròpies webs pel seu propi posicionament(ranking). Google es

partidària de l'ús de les dades estructurades i dels microformats. Prova d'això, serien les eines d'ajuda com l'assistent de marcatge de dades estructurades (*"Estructured Data Testing Tool"*). També, Google facilita una llista de com utilitzar els microformats, la qual tenen en compte alhora d'indexar i posicionar resultats de les cerques...

També, caldria destacar que, cada dia aquesta tècnica és més coneguda i implementada per desenvolupadors i, per aquesta raó compten amb eines de validació i creació d'aquests ítems dins del codi.

RDFas:

El seu objectiu es expressar el significat de les estructures del XML.

Tot sistema RDF parteix de tres entitats lògiques (triplets):

- Subjecte.
- Predicat.
- Objecte.

Això, fa que puguin suportar múltiples ítems per pàgina.

El **RDFa** (*Resource Description Framework in attributes*), afegeix extensions com a atribut, per poder incrustar metadades estructurades dins dels documents web. Els RDFa utilitzen les **etiquetes XHTML** per definir les propietats de l'objecte/persona. També, compta amb el suport del **W3C**.

L'exemple, a baix reflectit, ens mostra una manera complexa de programar en RDFa. Tot i així, aquesta manera de programar, aporta una millor extensibilitat al codi, permetent desenvolupar uns estàndards propis adaptats a les nostres necessitats.

```
<div xmlns:v="http://schema.org/" typeof="v:Product">  
   <!-- Ficticia -->  
  <span property="v:name">Exemple de producte</span>  
</div>
```

[Fig.29: Exemple complex d'una codificació en RDFa.](#)

Degut a la bona acceptació que van rebre les microdades, per part de Google i Schema.org, va aparèixer **RDFa lite** com alternativa a aquesta. Aquesta és un *"subset"* (subconjunt/derivació) del RDFa. Aquest, disposa d'una sintaxis molt més simple, que de fet, és pràcticament idèntica que les microdades, però aquesta, disposa d'extensibilitat.

Com a exemple, podria ser:

```
<div vocab="http://schema.org/" typeof="Product">  
    
  <span property="name">Marca Volkswagen: Tiguan.</span>  
</div>
```

[Fig.30: Execució del codi HTML amb RDFa.](#)

Com es pot veure, en exemples senzills, les microdades i els RDFas lite són quasi iguals. Com més complexa sigui el codi, més diferències es veuran, en:

- **RDFa lite**, no precisa de la sentència “*Itemref*”, utilitzada per les microdades per a relacionar diversos objectes.
- El llenguatge de les **Microdades**, no suporta la sentència “*prefix*”, utilitzada per barrejar diferents vocabularis en la mateixa web/document (Facebook, Google, etc).
- Les **microdades** no admeten la sentència “*vocab*”, utilitzada per especificar un vocabulari utilitzat en el “*markup*”.

En resum, l’ utilització de l’etiquetatge semàntic i d’Schema.org hauria de millorar la visibilitat de la pàgina web, facilitar que els cercadors la trobin fàcilment i, segons Google, crear Rich Snippets en els resultats de les cerques...tot això, hauria de facilitar la seva “visibilitat”...

També, s’observa l’elecció que tres motors de cerca d’una única sintaxis, les microdades, preferien aquesta a d’altres sintaxis com, RDFa. Aquesta última, essent molt més difosa i utilitzada per webmasters. La RDFa podria tenir les característiques per poder ser aplicada a Schema.org... Totes aquestes raons, ens donen lloc a algunes preguntes:

- L’etiquetatge en Schema.org afavoreix a millorar la “visibilitat” de la pròpia pàgina web? I que es trobi més ràpida? Quina es la més gran avantatge que té?
- Si s’utilitza el vocabulari d’ Schema.org, quina és la sintaxis, entre les microdades i el RDFa, que doni més bons resultats en el posicionament de la pròpia pàgina web?
- Com sabem Google va canviant i, amb ell, els seus resultats (amb el temps),...com pot afavorir l’actualització del contingut de la web al seu posicionament, dins de les demés webs?

2.5 Estudis amb Microdades, Microformats i RDFas, en Schema.org i OGP:

Simbiosi entre els diferents vocabularis(Schema,ogp) i les diverses sintaxis (microdades, microformats i RDFas):

Partint de la base que l'OpenGraphProtocol i l'Schema.org són dos vocabularis d'una llarga llista com; Dublin Core, SIOC o FOAC...

L'Schema.org i l'OGP no disposen d'una sintaxi específica. Per tant, si es vol implementar, es pot utilitzar la sintaxi de les microdades, microformats o els RDFas. Si s'utilitza el vocabulari de l'Schema.org, es podrà utilitzar qualsevol de les tres sintaxis. En canvi, si és implementar la web amb referències de xarxes socials, s'haurà d'implementar en vocabulari Open Graph Protocol i amb metadades de sintaxi, dins del mateix codi de la web.

També, penso que s'ha de tenir molt clar que per l'estructuració i contextualització de la mateixa web, pot ser necessari d'incloure més d'un vocabulari i/o més d'una sintaxi, a la vegada. Cada una de les sintaxis s'encarregarien d'estructurar aquella part d'informació per a contextualitzar-la millor dins de la mateixa web. Tot això, es fa a partir de la necessitat que tingui la mateixa web i perquè els diferents cercadors, puguin entendre millor el significat real de la informació de la web. Com més estructurada, definida, compatible i contextualitzada tinguem la informació, més fàcil serà que els cercadors la reconguin, tardin menys temps en analitzar-la i per tant, es valori positivament dins dels algorismes de cerca. Originant que aquests ho valorin –en certa manera- dins del rànquing del posicionament de la mateixa web...

```
<!-- namespace declaration -->
<html prefix="og: http://ogp.me/ns#">
  <!-- define microdata scope and type -->
  <head itemscope itemtype="http://schema.org/Article">
    <title>Social Site Example</title>
    <!-- define ogp and itemprop of microdata in one line -->
    <meta property="og:title" itemprop="name" content="Enjoy Fireworks">
    <!-- define ogp image -->
    <meta property="og:image" content=
      "https://developers.google.com/web/imgs/fireworks.jpg">
    <!-- use link[href] to define image url for microdata -->
    <link itemprop="image" href="//developers.google.com/web/imgs/fireworks.jpg">
    <!-- define ogp and itemprop of microdata in one line -->
    <meta property="og:url" content=
      "discovery-and-distribution/optimizations-for-crawlers/social-sites2.html">
    <!-- define ogp type -->
    <meta property="og:type" content="website">
    <!-- define twitter cards type -->
    <meta name="twitter:card" content="summary_large_image">
    <!-- define site's owner twitter id -->
    <meta name="twitter:site" content="agektmr">
    <!-- define description for ogp and itemprop of microdata in one line -->
    <meta property="og:description" itemprop="description"
      content="Fireworks are beautiful. This article explains how beautiful fireworks are.">
    <!-- general description (separate with ogp and microdata) -->
    <meta name="description" content="Fireworks are beautiful.
      This article explains how beautiful fireworks are.">
  </head>
```

Fig.31: Exemple d'inserció d'Schema i OGP en la mateixa web.

L'exemple de dalt, es basa en una web, la qual necessita poder crear fragments enriquits per a què el motor de cerca Google+ pugui analitzar amb facilitat i comprendre el seu contingut. Gràcies al vocabulari Schema.org i les microdades, ajuden a què els motors de cerca entenguin millor el contingut de la pàgina. S'ha de destacar que, mentre que la majoria de les metadades es troben incrustades dins de la capçalera, les microdades es troben dins del mateix context. L'OpenGraphProtocol, proporciona Facebook amb les metadades necessàries per garantir que les mateixes pàgines web tinguin la mateixa funcionalitat que altres objectes del Facebook, dins d'ell. Si les metadades s'inclouen dins de la capçalera de la web, aquestes s'utilitzen per poder obtenir informació de fragments enriquits quan es comparteix la pàgina. També cal destacar que, les etiquetes "meta" proporcionen la informació semàntica als mateixos rastrejadors dels llocs socials (Facebook i Goggle+).

Cap al final del codi de dalt, s'insereixen les "Twitter Cards", aquestes són una extensió del OGP però aquest, aplicable a Twitter. Els permet afegir arxius adjunts multimèdia als mateixos tuits, mitjançant enllaços a la seva pàgina web i, amb la riquesa de detalls que s'hagin afegit gràcies a la twitter card. També, en assignar l'ID de "twitter:site", aquest incorpora la informació en el post compartit per a què qualsevol persona pugui interactuar fàcilment amb el propietari de la pàgina.

Estudis realitzats per WebDataCommons:

WebDataCommons va realitzar dos estudis dins d'una franja de temps molt específica, d'un any. La franja de temps, anava des del Desembre del 2014 fins al Novembre del 2015. Les pàgines webs que es van rastrejar en un estudi i en l'altre son –quasi bé- iguals (la pobra variació és inapreciable). Per poder fer l'estudi, es van tenir en compte tres factors, les tres sintaxis més importants en aquella època; Microdades, Microformats i els RDFas. Crec que, mirant els resultats obtinguts les dues vegades en què es va fer la prova i en un marge de temps no massa llarg, podem observar les tendències que té la gent alhora d'etiquetar i estructurar les dades i en quins dominis utilitzen més... No fa falta esmentar que, per defecte, podem arribar a concloure si el que hem dit fins ara –en el TFM- és correcte o no (basant-nos sempre en els resultats obtinguts en aquests dos estudis)...

✚ Si observem les dades obtingudes durant el **Desembre del 2014**:

Overall

Crawl Date	Winter 2014
Total Data	160 Terabyte (compressed)
Parsed HTML URLs	2,014,175,679
URLs with Triples	620,151,400
Domains in Crawl	15,668,667
Domains with Triples	2,722,425
Typed Entities	5,516,068,263
Triples	20,484,755,485

[Fig.32- Rastreig comú realitzat al Desembre del 2014 per webdatacommons: Totals](#)

Segons WebDataCommons, durant el mes de Desembre del 2014 es va efectuar el rastreig dins la xarxa, obtenint un total de 160 TB d'informació comprimida. Durant el rastreig, es van analitzar 2.014.175.679 urls en HTML. D'aquestes, unes 620.151.400 urls, contenien Triples. També, es van rastrejar 15.668.667 dominis, dels quals, 2.722.425 amb Triples.

Results per Format

Format	Domains	URLs	Typed Entities	Triples
html-rdfa	571,581	257,251,367	405,541,283	2,566,827,347
html-microdata	819,990	292,601,824	2,209,497,281	9,438,536,906
html-mf-geo	20,261	4,619,664	7,712,041	20,348,760
html-mf-hcalendar	24,208	3,496,061	34,595,069	169,557,078
html-mf-hcard	1,095,517	101,606,009	1,349,620,300	3,850,290,103
html-mf-hcard including html-mf-adr	1,269,607	131,027,251	1,430,171,657	4,104,910,682
html-mf-hlisting	3,167	202,889	4,473,631	18,838,183
html-mf-hrecipe	3,476	630,402	5,781,217	24,756,234
html-mf-hresume	155	16,343	82,751	462,002
html-mf-hreview	13,772	2,496,303	16,186,868	69,802,632
html-mf-species	96	31,444	218,463	653,111
html-mf-xfn	170,202	17,032,646	52,187,702	219,772,447

[Fig.33- Rastreig comú realitzat al Desembre del 2014 per webdatacommons: Resultats per Format.](#)

En aquest apartat, només vull mencionar les quantitats per formats. Per aquesta raó, he cregut convenient englobar tots els resultats de microformats en un de sol. Per tant, obtenim les següents dades:

	Dominis
MICRODADES	819.990
MICROFORMATS	2.600.461
RDFa	571.581

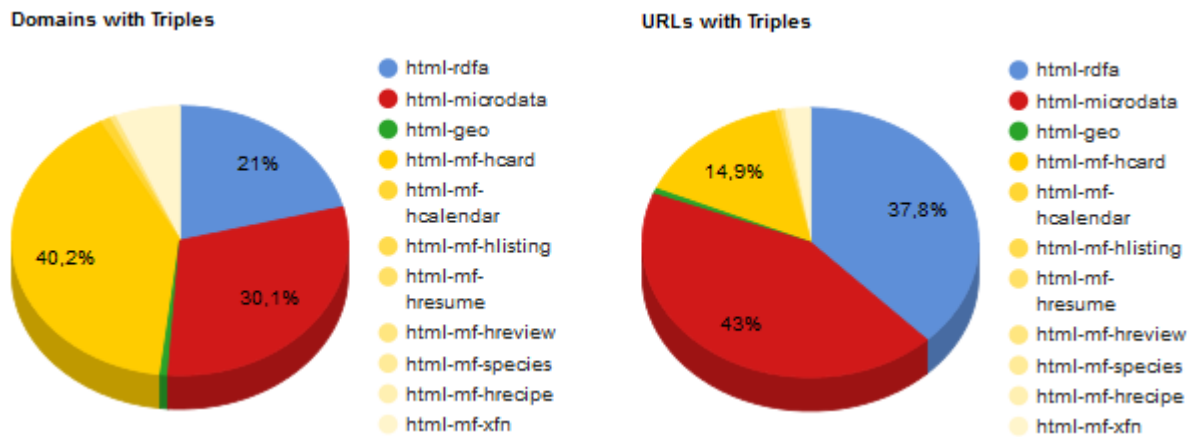


Fig.34- Rastreig comú realitzat al Desembre del 2014 per webdatacommons: Grafiques(I)

Podem observar que, mentre els dominis amb Triples estan formats –majoritàriament - per microformats (49%). Les URL amb Triples també marquen majoria però, en aquest cas, amb microdades (43%).

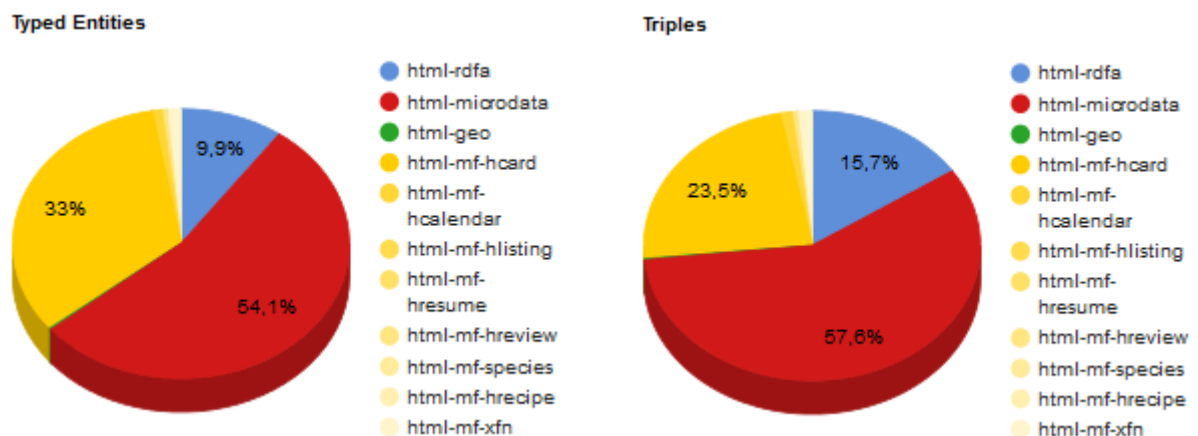


Fig.35- Rastreig comú realitzat al Desembre del 2014 per webdatacommons: Grafiques(II)

Els resultats obtinguts sobre les entitats tipades es poden observar en la figura de dalt. Mentre que clarament, la majoria s'estructuren amb microdades (54,1%), el segueix –amb un 36% del total- les estructurades amb microformats i, la resta, amb un pobre 10% s'estructuri amb RDFa. De la mateixa manera succeeix amb els Triples; la majoria s'estructuren amb microdades(57,6%) , seguidament dels microformats (26,7%) i per finalitzar, els RDFas (15,7%).

✚ Si observem les dades obtingudes el **Novembre del 2015**:

Segons l'estudi, realitzat al Novembre del 2015 per "WebDataCommons" (<http://webdatacommons.org>), ens dona a conèixer que 24,4 milions de pàgines web es troben estructurades amb; microdades, microformats o RDFas. De les 1770 milions de webs cercades per fer l'estudi, es van trobar 541 milions de pàgines web amb dades estructurades, representant un 30% de la recerca. Dins d'aquest 30%, trobem:

- ~521.000 utilitzen RDFa.
- 1.100.000 utilitzen Microdades.
- +1.000.000 utilitzen Microformats.

Overall

Crawl Date	November 2015
Total Data	151 Terabyte (compressed)
Parsed HTML URLs	1,770,525,212
URLs with Triples	541,514,775
Domains in Crawl	14,409,425
Domains with Triples	2,724,591
Typed Entities	6,107,584,968
Triples	24,377,132,352

[Fig.36- Rastreig comú realitzat al Novembre del 2015 per webdatacommons: Totals](#)

Segons WebDataCommons, durant el mes de Novembre del 2015 es va efectuar el rastreig dins la xarxa, obtenint un total de 151TB d'informació comprimida. Durant el rastreig, es van analitzar 1.770.525.212 urls en HTML. D'aquestes, unes 541.514.775 urls, contien Triples. També, es van rastrejar 14.409.425 dominis, dels quals, 2.724.591 amb Triples.

Results per Format

Format	Domains	URLs	Typed Entities	Triples
html-microdata	1,100,783	312,229,919	2,463,165,945	13,224,134,881
html-rdfa	521,806	196,336,975	460,658,558	1,598,114,462
html-embedded-jsonld	596,229	35,486,192	72,180,127	382,896,204
html-mf-adr	140,063	13,937,305	45,073,622	146,968,811
html-mf-geo	16,263	3,802,038	6,995,794	17,847,779
html-mf-hcalendar	25,721	3,007,502	24,846,049	119,168,060
html-mf-hcard	1,189,863	104,318,715	1,471,748,370	4,238,311,400
html-mf-hcard incl. html-mf-adr	1,198,702	105,712,446	1,516,821,992	4,385,280,211
html-mf-hlisting	2,484	135,780	4,043,961	16,126,218
html-mf-hrecipe	2,862	389,108	3,504,888	13,456,521
html-mf-hresume	115	2,987	6,741	19,338
html-mf-hreview	11,469	1,637,792	6,048,539	37,280,831
html-mf-species	88	21,370	138,818	382,344
html-mf-xfn	139,426	16,159,224	32,351,564	197,145,292
overall	2,724,591	541,514,775	6,107,584,968	24,377,132,352

[Fig.37- Rastreig comú realitzat al Novembre del 2015 per webdatacommons: Resultats per Format.](#)

En aquest apartat, només vull mencionar les quantitats per formats. Per aquesta raó, he cregut convenient englobar tots els resultats de microformats en un de sol i també, no tenir en compte el JSON-ld per poder realitzar una comparació realista amb les dades obtingudes del Desembre del 2014 i del Novembre del 2015. Per tant, obtenim les següents dades:

	2015
MICRODADES	1.100.783
MICROFORMATS	2.727.056
RDFa	521.806

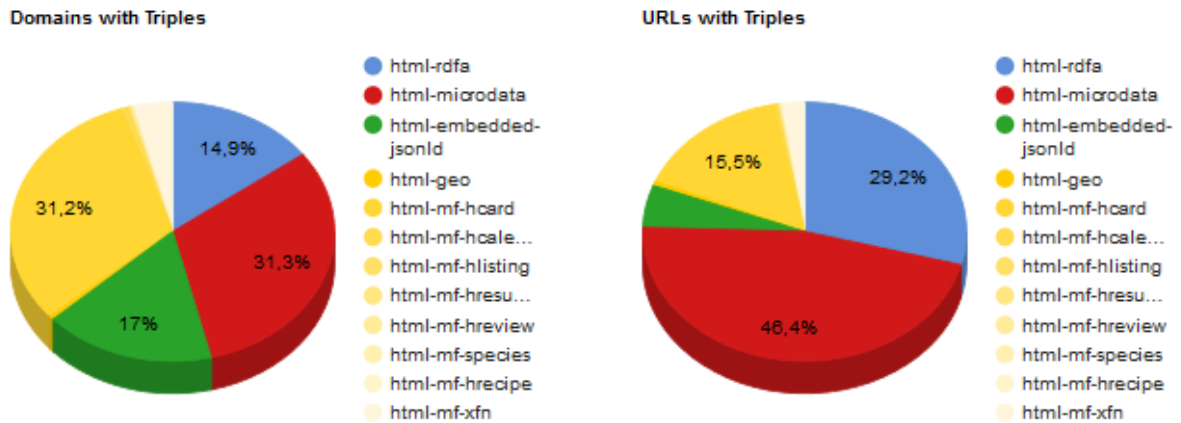


Fig.38- Rastreig comú realitzat al Novembre del 2015 per webdatacommons: Grafiques(I)

Podem observar que, mentre els dominis amb Triples es realitzen tant en microformats(35,7%) com en microdades(31,3%), majoritàriament. En canvi, les URL amb Triples marquen majoria, en aquest cas, amb microdades (45,4%), seguidament del RDFa (29,2%).

Si comparem aquestes gràfiques amb les del Desembre del 2014, tenim que; els dominis amb triples, té un descens de l'11% en els microformats i del 7% en els RDFas. En canvi, en les microdades el percentatge no pateix variacions. Aquest ~18% de descens que pateixen els microformats i els RDFas, és el percentatge que s'utilitza en la sintaxi JSON-ld durant el mes de Novembre. Pel que fa a les urls amb triples, podem observar que els microformats no pateixen cap variació. En canvi, les RDFas tenen un descens del ~9%. Al contrari que les anteriors, les microdades tenen una pujada significativa del 3,4%.

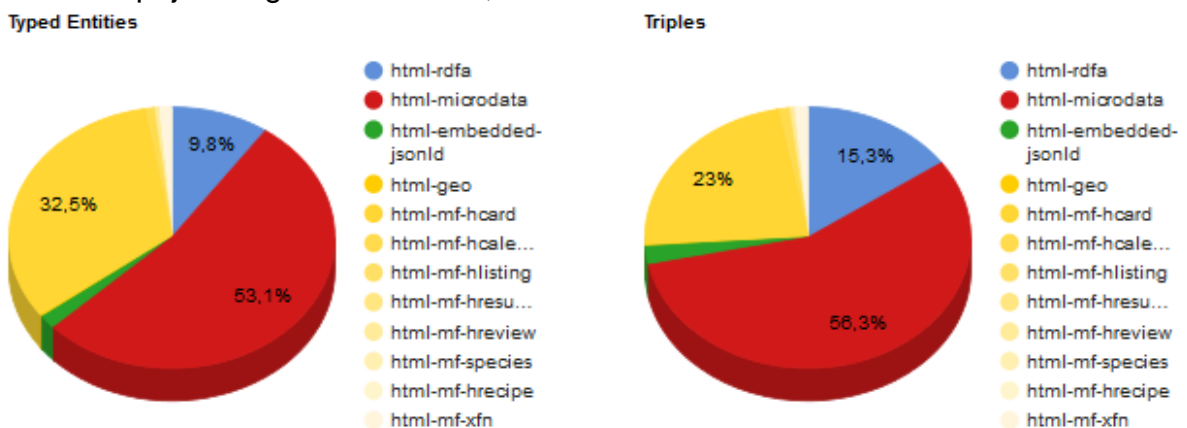


Fig.39- Rastreig comú realitzat al Novembre del 2015 per webdatacommons: Grafiques(II)

Els resultats obtinguts sobre les entitats tipades es poden observar en la figura de dalt. Mentre que clarament, la majoria s'estructuren amb microdades (53,1%), el segueix –amb un 35% del total- les estructurades amb microformats i, la resta, amb

un pobre 10% s'estructuren amb RDFa. De la mateixa manera succeeix amb els Triples; la majoria s'estructuren amb microdades(57,6%) , seguidament dels microformats (25,8%) i per finalitzar, els RDFas (15,7%).

Si comparem aquestes gràfiques amb les del Desembre del 2014, tenim que; no es percep cap variació significativa en les entitats tipades ni en els triples. Per tant, això ens indica que les persones encara utilitzen les microdades per crear les entitats i crear les relacions triples. Val a dir que, les microdades no són llegibles pels analitzadors de dades enllaçades, s'han desenvolupat eines per a poder agafar les microdades des del HTML. Un cop absorbides, es realitza la seva conversió al model RDF. Per aquesta raó, la gent utilitza més la sintaxi de les microdades, ja que es més fàcil d'utilitzar i, si a més, es fa una conversió d'aquestes al model RDF per poder tractar les dades, per què utilitzar el RDFa?

 **Conclusions**, a partir dels resultats de l'estudi:

	2014	2015	UTILITZACIÓ
MICRODADES	819.990	1.100.783	AUGMENT DEL 34,24%
MICROFORMATS	2.600.461	2.727.056	ESTABLE
RDFa	571.581	521.806	ESTABLE

Podem veure l'increment **(+34%)** que ha tingut les **microdades** en l'etiquetatge de les webs. Cada vegada més, les persones tendeixen a utilitzar allò que és més fàcil de fer anar i d'entendre. Tanmateix, influencia força que aquests, sigui la sintaxi del marcatge semàntic recomanada per **Schema.org**, ja que, afavoreix la visibilitat de la web i l'accés al seu contingut. També influencia que, **Google** llegeixi el seu format de text enriquit...

Com es pot comprovar per les dades, l'evolució de les microdades és notable. Aquests, al ser un conjunt d'ítems (estructures de dades <nom,valor>;) i que s'incorporen al HTML5, com si fossin atributs descriptius del tipus d'identitat a la qual fan referència a la informació que conté la mateixa web. D'aquesta manera, se li faciliten les cerques a Google quan rastreja la web. El cercador podrà organitzar la informació de la web més ràpidament. Aquest factor, farà que el cercador tingui més en compte la web alhora de posicionar-la.

A més, em de tenir en compte que l'HTML5 pot oferir serveis (de tot tipus) en diferents dispositius que no admetin la instal·lació de complements de tercers(p.ex. mòbils).

2.6 Conclusions:

La utilització de l'etiquetatge semàntic, permet una millor estructuració de la informació i una profunda comprensió del significat, per part del cercador. També, un factor a tenir en compte és, que la seva utilització incrementa fins a un 30% la visibilitat de la mateixa informació, i per tant, de la mateixa web. Això no és degut al que pensa la gent, que "és gràcies al mateix etiquetatge semàntic". La veritat és que és gràcies a les entrades (links) que té la mateixa web per poder veure la informació. Per poder tenir força entrades, la web ha de ser; "atractiva", "vistosa", "interessant" als ulls dels internautes per a què aquests, tinguin ganes d'entrar dins la web... Com més entrades tingui, més pujarà en el posicionament del rànquing. Per tant, sí que hi juga un paper important l'etiquetatge semàntic però, si la web no és "atractiva", poca cosa es podrà fer...

Des del meu humil punt de vista, hi ha dos factors a tenir en compte alhora de determinar quina metodologia de marcatge utilitzar:

Si el marcatge s'ha d'utilitzar en **pàgines web**, la millor manera de fer-ho és amb l'etiquetatge d'**Schema.org** i el format de les **Microdades** introduïdes amb HTML5. S'ha tingut en compte que, Schema.org és l'únic sistema de marcatge de dades compatible amb els principals motors de cerca; Google, Yahoo, Bing, Ask, etc. A més, Schema.org anima a què s'utilitzin les microdades. Aquests poden ajudar a què el cercador entengui i mostri, de manera estructurada, la informació que té la web. Aquesta manera d'estructurar la informació es basa en l'etiquetatge per "nom-valor". La seva gran simplicitat en la seva implementació fa que sigui més senzill i entenedor utilitzar les microdades que el RDFa, ja que és més complex. La seva utilització és molt senzilla, intuïtiva i res intrusiva. A més, els podem utilitzar per multitud d'estructures compatibles amb Schema.org.

Si el marcatge s'ha d'utilitzar dins de les **xarxes socials**, la millor manera de fer-ho seria amb **OGP** i les seves **metadades**. És un mètode senzill que ens permet introduir metadades dins la nostra pàgina web, per poder-la convertir en un objecte "Social Graph" (ja que, al entrar dins d'una xarxa social, ens convertim en objectes). En el moment d'entrar, passes a formar part d'un mapa de connexions socials a gran escala, en aquest món s'interactua amb "likes"(Facebook) o "share"(Google+). Quan inserim OGP dins la nostra web, estem creant un objecte de la xarxa social el qual podem manipular.

En OGP no tenim les estructures jeràrquiques tan extenses com les de l'Schema.org. Només necessita deu metadades per poder estructurar la informació de la web, però aquesta no queda del tot ben definida. En canvi, en Schema.org, s'estructura molt bé la informació i, a més, facilita la creació d'altres contenidors per augmentar la seva conceptualització. Per poder-ho fer, necessita moltes més

etiquetes i codi que l'OGP.

Mentre que Google anima a utilitzar Schema.org, Facebook ha acceptat l'OGP per millorar la visibilitat i les publicacions externes del Facebook i twitter. L'OGP va molt bé perquè no és necessari gran volum de contextualització de les dades.

En definitiva, tant Schema.org com TheOpenGraphProtocol són importants dins del seu entorn:

- Si necessitem estructurar la informació en pàgines webs, és important utilitzar l'Schema.org i les Microdades.
- Si volem incloure la nostra web dins de les xarxes socials, és important utilitzar l'OGP per manipular les dades dins de la xarxa social.
- Si volem fer les dues coses, estructurar les dades i incloure la web dins d'una xarxa social, és important utilitzar les dues; Schema.org amb microdades (per l'estructuració de les dades), i l'OGP amb metadades (per crear l'objecte de la xarxa social i poder-la manipular).

En resum, tant Schema.org com OGP són molt importants i necessaris per a la web semàntica.

Bibliografia.

Pàgines web:

Web Semàntica: <http://semanticweb.org>

Web oficial Schema.org: <http://schema.org>

Organització de les estructures d'Schema.org: <http://Schema.org/docs/schemas.html>

Web oficial The Open Graph Protocol: <http://ogp.me>

Web oficial W3C: <http://www.w3.org>

W3C Guia de Web Semàntica:

<http://www.w3c.es/divulgacion/guiasbreves/WebSemantica>

Microdades:

<https://www.w3.org/TR/microdata/>

<https://developers.google.com/schemas/formats/microdata>

Microformats:

<http://microformats.org>

<https://es.wikipedia.org/wiki/Microformato>

RDFa:

<https://www.w3.org/TR/rdfa-lite/>

<https://es.wikipedia.org/wiki/RDFa>

Articles i altres documents:

Estudi realitzat per WebDataCommons(Desembre, 2014):

<http://webdatacommons.org/structureddata/2014-12/stats/stats.html>

Estudi realitzat per WebDataCommons(Novembre, 2015):

<http://webdatacommons.org/structureddata/2015-11/stats/stats.html>

Eines:

Structured Data Testing Tool:

<https://developers.google.com/structured-data/testing-tool/>

Webmaster Tools de Google: <https://www.google.com/webmaster/tools/home?hl=ca>

Facebook for developers: <https://developers.facebook.com/tools/debug>