

Estudi de la fiabilitat de la Viquipèdia com a recurs terminològic

Antoni Oliver
Mercè Vázquez
Georgina Ubide



Antoni Oliver
Universitat Oberta de Catalunya
aoliverg@uoc.edu;
ORCID:
[0000-0001-8399-3770](https://orcid.org/0000-0001-8399-3770)



Mercè Vázquez
Universitat Oberta de Catalunya
mvazquezga@uoc.edu;
ORCID:
[0000-0002-7983-4029](https://orcid.org/0000-0002-7983-4029)



Georgina Ubide
Universitat Oberta de Catalunya
gubide@uoc.edu;
ORCID:
[0000-0003-4825-2285](https://orcid.org/0000-0003-4825-2285)

Resum

En l'àmbit de la traducció especialitzada es considera que la Viquipèdia no és un recurs d'informació especialitzat fiable degut al fet que qualsevol usuari, sigui especialista o no de la matèria, pot redactar una entrada. En aquest article es vol determinar el grau de fiabilitat de la Viquipèdia com a recurs terminològic especialitzat per als traductors, a partir d'unes tasques que hem definit per tal d'avaluar i valorar el contingut especialitzat.

Paraules clau: Viquipèdia; terminologia; àmbit especialitzat; extracció de termes; candidats a terme

Resumen

En el ámbito de la traducción especializada se considera que la Wikipedia no es un recurso de información especializada fiable debido a que cualquier usuario, sea especialista o no de la materia, puede redactar una entrada. En este artículo se pretende determinar el grado de fiabilidad de la Wikipedia como recurso terminológico especializado para los traductores, a partir de unas tareas que hemos definido con el fin de evaluar y valorar el contenido especializado.

Palabras clave: Wikipedia; terminología; ámbito especializado; extracción de términos; candidatos a término

Abstract

In Translation Studies it is often said that Wikipedia is not a reliable source for specialized content, because it can be written by either experts or non-experts. This study aims to demonstrate Wikipedia's reliability as a source of terminology for translators. We have established a range of methods to assess and evaluate specialized content.

Keywords: Wikipedia; terminology; specialised domain; term extraction; term candidates

Revista Tradumàtica. Número 15 (desembre 2017).

Cita recomanada: OLIVER, A.; VÁZQUEZ, M.; UBIDE, G. (2017). «Estudi de la fiabilitat de la Viquipèdia com a recurs terminològic». Revista Tradumàtica. Tecnologies de la Traducció, 15, 10-20.
<https://doi.org/10.5565/rev/tradumatica.193>



<http://revistes.uab.cat/tradumatica>

ISSN 1578-7559

Agraïments: Per a fer la revisió dels termes de l'àmbit de dret extrets de la Viquipèdia hem comptat amb la col·laboració de sis professors dels Estudis de Dret de la Universitat Oberta de Catalunya. Gràcies al seu interès i a la seva implicació hem pogut assegurar la revisió que hem reatzenat en terminologia especialitzada.

1. Introducció

Durant les últimes dècades, gràcies a l'augment de l'accessibilitat a Internet, la xarxa mundial viu un canvi significatiu a mesura que cada cop més i més persones contribueixen activament al contingut que hi ha disponible. Les tecnologies han permès i permeten la compilació de dades de tots els àmbits d'estudi, i la necessitat d'agrupar el coneixement per àmbits ha impulsat d'un temps ençà l'extracció de terminologia dels diferents camps de coneixement. Aquesta tasca és interessant i enriquidora tant per a lingüistes, traductors, intèrprets, terminòlegs, i en general per a professionals especialitzats, com per a futurs professionals i estudiants. Les aplicacions que tenen les terminologies obtingudes són diverses, entre les quals destaquen:

- *L'elaboració de glossaris i diccionaris terminològics, que seran una referència per a qualsevol persona que escrigui sobre un tema específic.*
- *L'elaboració de diccionaris terminològics multilingües, que faciliten i agilitzen la traducció automàtica, ja que contribueixen a la detecció de termes específics i l'aclariment de polisèmies.*
- *La construcció d'ontologies de dominis semàntics concrets.*

Una de les majors i més populars obres de consulta que podem trobar actualment en línia és la Viquipèdia. Es tracta d'una enciclopèdia lliure, multilingüe i editada de manera col·laborativa, creada el 2001 amb l'objectiu de generar i distribuir una enciclopèdia lliure d'alta qualitat perquè qualsevol persona del planeta pugui accedir al coneixement en el seu idioma. Ara per ara, l'obra consta de més de 37 milions d'articles en 287 idiomes. La raó d'aquest augment tan accelerat de disponibilitat de coneixement es deu al fet que es redacta de manera conjunta per voluntaris arreu del món. Com que es tracta d'un *wiki*, pot ser editada per qualsevol usuari. Aquest tret característic de l'enciclopèdia, pensat perquè qualsevol usuari posi a disposició lliure més informació o actualitzi el contingut disponible, ha contribuït a les crítiques negatives que rep l'obra. No obstant, de vegades ens trobem que la consulta feta sí que és útil, o que altres llocs de referència, com la base de dades terminològica de la Unió Europea (IATE¹), ens redirigeixen a aquesta enciclopèdia.

En la part que ocupa als traductors, és important i necessari utilitzar recursos d'informació i de coneixement fiables per a fer consultes, però també per a poder crear eines pròpies que ajudin i assegurin la seva feina. Com demostra Alonso (2015), els traductors professionals tenen una bona opinió de la Viquipèdia, concretament en tres aspectes: utilitat, fiabilitat i facilitat d'ús. El seu estudi es basa en enquestes i analitza aspectes que estan relacionats amb el nostre estudi, com ara trobar el significat d'un terme, localitzar termes equivalents i crear glossaris propis. Per aquest motiu, en aquest treball es vol determinar el grau de fiabilitat d'aquesta gran enciclopèdia com a recurs terminològic especialitzat per als traductors per a crear alguna eina de suport (bases de dades terminològiques, diccionaris, glossaris).

¹ IATE: <<http://iate.europa.eu/>> [Consulta: 18 de novembre de 2017]

A continuació es presenta el marc teòric del treball, la metodologia i els recursos aplicats per a dur a terme l'estudi; es mostren i s'analitzen els resultats obtinguts; i finalment s'inclouen les conclusions que se'n formulen.

2. Marc teòric

2.1 La terminologia

El terme es defineix com una unitat lèxica que expressa un concepte en una àrea temàtica específica. El problema principal és la detecció d'aquests termes, ja que els extractors ofereixen termes candidats, en lloc de termes pròpiament dits. Tal i com apunten Vivaldi i Rodríguez (2011), els termes han de complir unes propietats com ara la coherència interna (*unithood*), relació entre terme i concepte (*termhood*), i el seu ús especialitzat.

Així mateix, Pazienza (2005) defineix un terme com "una representació superficial d'un concepte en un domini específic", entenent com a representació superficial la denominació i per domini específic el camp d'especialitat. Sovint els traductors no tenen en compte les teories i definicions sobre terminologia i inclouen en les seves recopilacions terminològiques unitats que no poden ser considerades com a termes però que demanen una gran consistència en la seva traducció (Oliver, 2016). En aquest treball, però, tindrem en compte la definició més estricta d'unitat terminològica per a valorar el nivell de fiabilitat de la Viquipèdia com a recurs terminològic.

Generalment, les terminologies s'obtenen a partir de corpus textuais on es poden trobar els termes en actiu fent ús d'eines semiautomàtiques per a la detecció d'unitats terminològiques. D'acord amb Medelyan et al. (2009), aquests recursos especialitzats ofereixen una alta qualitat de contingut però es veuen restringits en volum i cobertura. Són els experts qui els redacten i és impossible que coneguin tots els temes i els descobriments que van sorgint. Per això l'actualització d'aquests continguts no és constant.

2.2 La Viquipèdia

La Viquipèdia presenta unes característiques combinades entre alta qualitat i alta cobertura: tots els articles conformen una base de coneixement entrelaçat reforçada amb un sistema de categorització i etiquetatge, que donen aquesta propietat de corpus a l'enciclopèdia. Des de la seva creació, ha inspirat un gran nombre d'estudis en lingüística i informàtica. Per a dur a terme aquest projecte, els estudis de recerca de funcionament (Cajal, 2014) i d'aprofitament de la informació pels processadors del llenguatge natural (Zesch, Müller, i Gurevych, 2008), i estudis d'obtenció de terminologia temàtica tant monolingüe (Vivaldi i Rodríguez, 2012; Vivaldi i Rodríguez, 2014) com bilingüe (Erdmann et al., 2008) han inspirat l'enfocament del treball.

L'enciclopèdia està formada per un gran nombre d'articles classificats mitjançant un sistema de categories i d'etiquetes que permet agrupar i estructurar el contingut dels articles per àmbits de coneixement. Com a resultat d'aplicació d'aquest sistema, els articles es troben relacionats i vinculats a partir d'aquest etiquetatge, formant una gran xarxa d'informació. El problema d'aquest sistema, però, se centra en dos punts fonamentals:

- 1/ No hi ha consens global en la classificació de les disciplines acadèmiques, pel fet que és molt difícil marcar la línia que delimita un àmbit d'un altre.
- 2/ Els articles els pot redactar qualsevol usuari; per tant, depenent del contingut i del criteri de l'autor, es classifiquen segons les categories que més s'adeqüen a la informació.

Els articles, doncs, estan categoritzats a partir d'uns criteris subjectius que poden no coincidir, de manera general, amb les categories establertes. Tot i així, quan s'efectua una extracció mitjançant una categoria determinada, s'obtenen totes aquelles entrades classificades en aquesta categoria, siguin pertinents o no realment d'aquell àmbit concret.

3. Metodologia

Amb la investigació que es presenta es vol determinar el grau de fiabilitat de la Viquipèdia com a recurs especialitzat per a traductors, amb l'objectiu d'aprofitar el seu contingut com a eina de reforç (glossaris, diccionaris) que agiliti i asseguri la tasca traductora. Com a llengua principal d'estudi s'ha escollit l'anglès pel fet de ser actualment la llengua franca de la ciència i de la comunicació internacional, de manera que el volum d'informació disponible en aquesta llengua és més elevat. Quant al domini que s'avalua, s'ha escollit l'àmbit el dret.

L'accés a la Viquipèdia s'ha fet a través del projecte dBpedia², que ofereix un accés al contingut de l'enciclopèdia (tant l'anglesa com en algunes altres llengües) en una estructura del tipus base de dades. La seva estructura fa que sigui molt més fàcil de tractar computacionalment que no pas els *dumps* directes de la Viquipèdia, que estan en un format XML que conté format *wiki* en el seu interior. Per contra, el contingut de la dBpedia no s'actualitza de manera tan freqüent com els *dumps* de la Viquipèdia. Per a realitzar aquesta investigació, s'han fet servir unes taules concretes de la dBpedia, amb les quals hem delimitat les tasques per a avaluar la precisió terminològica:

- *SKOS categories: ens proporciona informació sobre els conceptes que són categories i com les categories estan relacionades entre sí. L'ús d'aquesta taula ens permet realitzar la primera tasca, crear la llista de termes d'un àmbit determinat: prenem la categoria general que correspon a l'àmbit que volem analitzar i busquem les categories relacionades, i repetim el procés fins a un cert nivell, que en els nostres experiments ha estat de 3.*

² DBpedia <<http://wiki.dbpedia.org/>> [Consulta: 18 de novembre de 2017]

- *Interlanguage links: ens proporciona el títol de l'entrada equivalent en una altra llengua i aquesta informació la podem fer servir per a trobar la traducció del títol d'una entrada. L'ús d'aquesta taula ens permet realitzar la segona tasca, obtenir i revisar les traduccions a l'espanyol i al català.*
- *Short abstracts: ens proporcionen les primeres oracions (fins aproximadament 600 caràcters) de les entrades i es poden considerar definicions. L'ús d'aquesta taula ens permet realitzar la tercera tasca, examinar el contingut de les definicions.*

L'inconvenient de seleccionar categories d'àmbits que actuen com a paraigua i agrupen moltes altres disciplines, és que als conjunts obtinguts el soroll és elevat. Si a aquest fet s'hi afegeix la categorització subjectiva dels articles per part de l'usuari –que és lliure, permet crear categories noves i modificar les categories existents–, el soroll s'incrementa.

3.1 Creació de llistes de termes

En la Viquipèdia, partint del nom de la categoria *law* que designa el domini escollit per al treball, se seleccionen les subcategories que engloba fins al nivell 3. S'ha escollit aquest àmbit general per a garantir que s'obté el màxim de coneixement disponible en aquest domini. A continuació s'obtenen totes les pàgines d'aquesta categoria i subcategories per a configurar un conjunt inicial de termes candidats de l'àmbit, extrets de manera automàtica juntament amb les primeres línies de definició. El conjunt s'estructura en forma de llista i s'avalua en format de full de càlcul.

ANGLÈS	Conjunt inicial de termes
Dret	2.114

Taula 1. Conjunt total de termes obtinguts

La taula 1 mostra la quantitat de termes que s'han obtingut de l'àmbit. L'extracció s'ha realitzat seleccionant els articles en llengua anglesa de l'àmbit analitzat i que estan enllaçats amb les entrades en espanyol i/o en català. Un cop hem obtingut el conjunt, hem identificat unitats que no són termes específics de l'àmbit que analitzem i apliquem una revisió per a descartar tot el que no és coneixement especialitzat:

- *Es detecta un gran nombre de named entities (NE), unitats lèxiques que es refereixen a una entitat concreta com ara una persona (Margot Frank), un lloc (Dreros), una organització (Parliament of Scotland) o un producte (cat café). Com que aquestes paraules no expressen coneixement especialitzat i tampoc no formen part de l'objecte d'estudi d'aquest treball, es marquen per a eliminar-les de la llista de termes.*
- *S'observa la presència de termes no pertinents a l'àmbit (NP), unitats lèxiques obtingudes per les categories que s'han seleccionat, però que no són de l'àmbit*

(public sector o ecocide). Aquestes paraules també es marquen per a eliminar-les de la llista de termes.

Les *named entities*, com a paraules no especialitzades, no formen part de l'objecte d'aquest estudi i no es poden tenir en compte per a valorar la precisió de coneixement especialitzat. Per tant, tot i que les hàgim obtingut en l'extracció, les hem de treure del conjunt per tal de calcular la precisió. Per a poder detectar les *named entities* del conjunt, es pot realitzar de manera automàtica mitjançant el sistema de les Infoboxes dels articles, és a dir, les taules amb la informació principal. Aquest sistema garanteix una detecció precisa d'aquestes entitats, però no tots els articles de la Viquipèdia presenten Infobox, i per tant, el sistema automàtic no les detecta. Per aquest motiu, la detecció de les NE s'ha realitzat de manera manual.

Per tal de garantir la revisió terminològica, hem demanat la col·laboració dels experts en aquest domini de coneixement per a confirmar els resultats que hem obtingut: El grup d'experts està format per professors dels Estudis de Dret de la Universitat Oberta de Catalunya (un total de sis professors han participat en la revisió dels termes). Per tal de poder dur a terme aquesta tasca, els professors han tingut en consideració que un terme es considera correcte si és un terme propi de l'àmbit de dret, i incorrecte tant si no és un terme com si es tracta d'un terme d'una altra especialitat.

- *Hi ha un total de 106 termes que són incorrectes (empowerment, ius singulare o capital punishment). Aquests termes els eliminem del conjunt.*
- *Del total de termes marcats per descartar, s'han mantingut al conjunt un total de 28 (laissez-faire, basic income o habeas data).*
- *D'un total de 24 termes en dubte de pertinença a l'àmbit, els experts han considerat que 10 són de l'àmbit (lex loci, lex fori o cause célèbre).*

ANGLÈS	NE	NP
Dret	862	717

Taula 2. Termes descartats

Després d'haver fet una primera revisió lingüística de les entrades i la revisió d'aquestes entrades per part dels experts, la taula 2 mostra la quantitat de *named entities* (NE) i de termes no pertinents (NP) trobats al conjunt. Un cop s'han detectat i eliminat, ens ha permès saber quants termes són realment correctes i d'aquesta manera poder calcular la precisió.

ANGLÈS	Nre. Termes correctes	Conjunt inicial NE	Precisió %
Dret	535	1252	42.73

Taula 3. Precisió sense NE

3.2 Trobar equivalències en espanyol i en català

Un cop tenim el conjunt de termes correctes en anglès, s'avaluen les seves equivalències en espanyol i català. En aquest cas, s'ha pogut calcular la precisió terminològica de les equivalències en espanyol i en català, així com la cobertura que ofereixen. No obstant, no parlem de cobertura en un sentit global, sinó de la quantitat de terminologia especialitzada que l'anglès té vinculada amb l'espanyol i amb el català.

	ESPANYOL			CATALÀ		
	Nre. termes	Precisió %	Cobertura %	Nre. termes	Precisió %	Cobertura %
Dret	502	98.04	95.70	271	98.90	51.21

Taula 4. Taula de precisió i cobertura de l'espanyol i el català a la Viquipèdia segons els termes correctes en anglès.

La taula 4 il·lustra el nombre d'equivalències de termes trobades en aquest àmbit en espanyol i en català. La revisió de les correspondències s'ha realitzat manualment i ha permès saber si els termes són correctes o no (precisió) i la quantitat de termes disponibles en aquests idiomes segons els termes correctes en anglès (cobertura). Per a identificar la precisió terminològica en aquests idiomes s'ha calculat quants termes són correctes en espanyol i en català, i la xifra obtinguda s'ha dividit entre el nombre de termes correctes en anglès. Per a calcular la cobertura s'ha dividit el nombre de termes en espanyol i en català entre el total de termes correctes en anglès.

3.3 Definicions dels termes

A banda de valorar si els termes d'aquest àmbit i en els idiomes escollits són correctes, s'ha realitzat també una revisió de les definicions que els acompanyen per a avaluar, d'una manera molt general, el contingut de coneixement de les pàgines de la Viquipèdia.

Les pàgines de l'enciclopèdia segueixen generalment el patró de diccionari enciclopèdic. Segons defineix el Centre de Terminologia TERMCAT³, aquests diccionaris “exposen el coneixement actual sobre les diverses matèries [...]. En aquests diccionaris no només s'explica el significat de les paraules, sinó que sobretot s'utilitzen aquestes entrades com a portes per a endinsar-se en una matèria determinada.” És a dir, contenen una primera part lexicogràfica introductòria acompanyada d'una segona part, que pot ser força extensa, de contingut enciclopèdic que aporta informació històrica, tècnica, etc.

Com veiem a les entrades de la Viquipèdia, consten d'un resum inicial que presenta la pàgina en qüestió, que pot estar format per diversos paràgrafs. No obstant, en l'extracció que s'ha dut a terme s'han obtingut les primeres línies definitòries dels termes, és a dir, els *short abstracts*.

³ TERMCAT <www.termcat.cat> [Consulta: 18 de novembre de 2017]

Per a valorar si el contingut de les definicions expressa coneixement especialitzat d'una manera global i assequible per un usuari no expert, hem examinat les definicions dels termes correctes en anglès i de les equivalències trobades en espanyol i català a partir d'uns criteris que hem delimitat:

- *que designi una unitat significativa i formal, és a dir, a cada concepte li correspon una definició;*
- *que la definició no faci referència a un concepte no especialitzat;*
- *que la definició no faci referència a un concepte d'un altre àmbit;*
- *que la definició no es refereixi al concepte especialitzat amb puntualització de pertinença en algun sentit;*
- *que la definició sigui precisa.*

Un cop aplicats aquests criteris, classifiquem les definicions en tres tipus: correcta, quan de manera general ofereix coneixement especialitzat; incorrecta, quan compleix algun dels criteris creats; o absent (-), quan no s'ha obtingut la definició del terme. Aquesta classificació ens ha permès quantificar, d'una manera general, el contingut especialitzat dels àmbits que analitzem.

DRET	C	I	-
EN [535]	515 (96.26%)	19 (3.55%)	1 (0.18%)
ES [512]	486 (90.84%)	10 (1.95%)	16 (3.12%)
CA [274]	229 (83.57%)	8 (2.91%)	37 (13.50%)

Taula 5. Resultats per àmbit i llengua de les definicions; àmbit de dret

La taula 5 mostra la quantitat de definicions classificades a cada tipus a partir dels criteris aplicats. La taula agrupa les diferents llengües que analitzem, acompanyades del nombre de termes correctes del conjunt, dels quals n'hem examinat les definicions. Les xifres obtingudes per a cada llengua van acompanyades del percentatge que expressa la proporció vers el total.

4 Comentari i anàlisi de les dades

Un cop s'han realitzat les tasques inicials establertes, s'observen una sèrie de resultats clars i diferenciats a cada fase que cal mencionar de manera individual. Els aspectes més significatius que s'han trobat al llarg d'aquesta investigació es presenten a continuació en el mateix ordre que hem realitzat les diferents tasques.

4.1 Els termes

En referència al punt central d'aquest treball, en general s'observen unes dades terminològiques força positives. D'una banda tenim els conjunts de termes en anglès, que ofereixen uns resultats diferents per cada àmbit analitzat.

S'observa la disponibilitat d'un volum elevat de contingut, però hi ha un elevat nombre d'entrades classificades com a pròpies de l'àmbit de dret que corresponen a NE i a NP (862 i 717, respectivament). Després del filtratge de contingut no especialitzat, tot i ser un conjunt amb un nombre elevat de termes candidats, els termes pròpiament de dret corresponen a un 40% del total d'entrades inicialment seleccionades.

Per altra banda, tenim els resultats obtinguts de les equivalències terminològiques en espanyol i en català, corresponents als termes detectats com a correctes en anglès. Com es pot observar a la taula 4, la terminologia obtinguda en ambdós idiomes supera el 95% de precisió. En canvi, s'adverteix una gran diferència en la quantitat de coneixement vinculat entre l'anglès i l'espanyol i l'anglès i el català: mentre que en espanyol s'ofereix una cobertura superior al 90% de coneixement especialitzat de l'àmbit analitzat, en català només es presenta una cobertura entorn del 50%.

4.2 Les definicions

Quant a les definicions dels termes, com ja s'ha comentat al punt 3.3, els articles de la Viquipèdia presenten de manera general el patró de diccionari enciclopèdic, la presentació d'informació és gradual (general > específic), però no hi ha unes normes establertes de redacció de les definicions. En conseqüència, no es poden revisar les definicions amb els mateixos criteris que s'apliquen a obres de coneixement especialitzat creades per institucions oficials. Per això s'han creat uns criteris que busquen la referència única i correcta d'un concepte especialitzat d'una manera general i assequible per a un usuari no expert, a partir dels quals s'han calculat els resultats de la taula 5.

Si mirem la taula 5, s'observa que en l'àmbit analitzat s'obtenen molt bons resultats referents a l'explicació dels termes. En anglès i en espanyol les definicions s'adeqüen i compleixen l'objectiu de presentar coneixement especialitzat en més d'un 90% dels casos. Si ens fixem en el català, entorn del 80% de definicions són apropiades, però el nombre de definicions correctes és menor degut al fet que l'absència de definicions destaca en comparació amb els altres dos idiomes: a la taula s'observa la falta d'explicació dels termes en català superior al 10%.

4.3 Observacions globals

A partir d'aquests resultats, es desprenen les següents observacions:

- *El volum de contingut inicial enllaçat en les llengües i l'àmbit del treball es pot considerar adequats, tenint en compte que hem extret termes d'una categoria especialitzada però general; no obstant, s'haurà de realitzar necessàriament un filtratge acurat de contingut per a descartar tot aquell coneixement que no és especialitzat o que és propi d'un altre àmbit.*
- *Amb relació als enllaços interlingüístics de les pàgines de la Viquipèdia en anglès, observem que hi ha un 50% més de pàgines enllaçades a la Viquipèdia espanyola que a la Viquipèdia catalana.*
- *Pel que fa a la terminologia que hem obtingut en català, veiem que hi ha una mancança d'explicació de coneixement especialitzat en comparació amb les altres llengües del treball.*

Podem dir, doncs, que certament compleix la seva tasca de ser una enciclopèdia lliure d'accés públic al coneixement especialitzat. No obstant, s'observa que queda per endavant una gran labor de vinculació de coneixement entre diferents llengües; i en el cas del català, hi ha molt de coneixement pendent de posar a disposició pública i una proporció important d'aquell coneixement que ja hi trobem es pot millorar quant a expressió de conceptes.

5 Conclusions, valoracions i treball futur

En aquest treball s'ha volgut determinar la fiabilitat de la Viquipèdia com a recurs especialitzat per a traductors. A partir de les tasques d'avaluació que hem definit inicialment i que ens han portat a realitzar diverses revisions manuals, podem determinar

positivament que la Viquipèdia presenta, de manera general, una bona fiabilitat del coneixement i de la informació que ofereix.

Per tant, si un traductor vol realitzar una consulta puntual, pot adreçar-se directament al cercador de l'enciclopèdia i navegar per la seva xarxa; però si, altrament, vol crear una eina de reforç que l'ajudi en la seva tasca, durant l'extracció de contingut haurà de tenir present les característiques d'organització de la informació i aplicar les revisions oportunes que li permetin prescindir d'aquell coneixement obtingut però que no necessita.

Veiem que la Viquipèdia encara té aspectes per millorar en un futur proper: i) perfilar el sistema de detecció de coneixement no especialitzat; ii) afinar el sistema de categories de contingut; iii) augmentar la disponibilitat de coneixement en altres llengües que van adquirint rellevància i presència; i iv) incrementar els enllaços interlingüístics.

Referències bibliogràfiques

- Alonso, Elisa (2015). "Analysing the use and perception of Wikipedia in the professional context of translation". *JoSTrans: The Journal of Specialised Translation*, n. 23 (January), p. 89-116. <http://www.jostrans.org/issue23/art_alonso.php> [Consulta: 10 de septiembre de 2017].
- Cajal Mariñosa, Sergio (2014). *Extracción de una terminología multilingüe de Wikipedia*. Proyecto final de carrera. Facultat d'Informàtica de Barcelona (FIB); Universitat Politècnica de Catalunya UPC. Barcelona. Enginyeria Informàtica, Pla 2003. <<https://upcommons.upc.edu/handle/2099.1/21683?locale-attribute=es>> [Consulta: 22 de julio de 2017].
- Erdmann, Maïke [et al.] (2008). "Extraction of bilingual terminology from a multilingual web-based encyclopedia". *Journal of Information Processing*, v. 16, (July), p. 68-79. <<http://doi.org/10.2197/ipsjip.16.68>> [Consulta: 10 de septiembre de 2017].
- Medelyan, Olena [et al.] (2009). "Mining meaning from Wikipedia". *International Journal of Human-Computer Interactions*, v. 67, n. 9, p. 716-754. <<https://arxiv.org/ftp/arxiv/papers/0809/0809.4530.pdf>>, <Doi: [10.1016/j.ijhcs.2009.05.004](https://doi.org/10.1016/j.ijhcs.2009.05.004)> [Consulta: 22 de julio de 2017].
- Oliver González, Antoni (2016). *Herramientas tecnológicas para traductores*. 1.^a ed. Barcelona: Editorial UOC. (Manuales; 423), (Manuales. Lingüística y traducción).
- Pazienza, Maria Teresa; Pennacchiotti, Marco; Zanzotto, Fabio Massimo (2005). "Terminology extraction: an analysis of linguistic and statistical approaches". En: Sirmakessis, S. (ed.). *Knowledge mining: proceedings of the NEMIS 2004 Final conference*. Berlin [etc.]: Springer, p. 255-279. (*Studies in Funzziness and Soft Computing*; 185). <https://doi.org/10.1007/3-540-32394-5_20> [Consulta: 4 de agosto de 2017].
- Vivaldi, Jorge; Rodríguez, Horacio (2012). "Using Wikipedia for Domain Terms Extraction". En: *Workshop on the Creation, Harmonization and Application of Terminology Resources: Proceedings of CHAT 2012: The 2nd Workshop on the*

- Creation, Harmonization and Application of Terminology Resources*, p. 3-10.
<<http://upcommons.upc.edu/handle/2117/23104>> [Consulta: 22 de julio de 2017].
- (2011). "Extracting terminology from Wikipedia". *Procesamiento del Lenguaje Natural*, n. 47 (septiembre), p. 65-73.<<http://journal.sepln.org/sepln/ojs/ojs/index.php/pln/article/view/981/734>> [Consulta: 22 de septiembre de 2017].
- (2010). "Finding domain terms using Wikipedia". En: *Proceedings of the 7th LREC International Conference*, p. 386-393. <http://www.lrec-conf.org/proceedings/lrec2010/pdf/748_Paper.pdf> [Consulta: 22 de julio de 2017].
- Zesch, Torsten; Müller, Christof; Gurevych, Iryna (2008). "Extracting lexical semantic knowledge from Wikipedia and Wiktionary". En: *Proceedings of the 6th LREC International Conference*, p.1646-1652.
<<https://pdfs.semanticscholar.org/065a/29adca32f66c16005de3f48ebb3512c8baf1.pdf>> [Consulta: 20 de agosto de 2017].