



ISPRA

Istituto Superiore per la Protezione
e la Ricerca Ambientale

**METODI E STRUMENTI PER LA REALIZZAZIONE
DI UN DIZIONARIO TERMINOLOGICO
SULLA TUTELA DELLE ACQUE**

ING. FRANCESCA ROMOLI

TUTOR: PROF. ATTILIO COLAGROSSI

CO-TUTOR: DOTT.SSA ANNAMARIA RIZZO

Data	Firma Tutor	Firma Stagista	Firma Responsabile Servizio

ANNO 2008/2009

Indice

Abstract	5
Introduzione	6
Metodologia	8
1 Proposta per un dizionario terminologico specialistico sulla tutela delle acque	9
1.1 Le normative di riferimento sulla tutela delle acque	10
1.2 Pianificazione delle attività	13
1.3 Scelta dei domini di interesse	14
1.4 Analisi delle normative ed estrazione dei termini	14
1.5 Individuazione di un primo nucleo terminologico	16
1.6 Individuazione delle relazioni tra termini	17
1.7 Inserimento dei termini	19
2 Il database lessicale WordNet	19
2.1 Genesi di WordNet	20
2.2 Forme di parola e significati	21
2.3 WordNet: organizzazione e principi di funzionamento	22
2.4 WordNet e le categorie lessicali	27
2.4.1 I nomi	28
2.4.2 Gli aggettivi	30
2.4.3 I verbi	32
2.4.4 Gli avverbi	33
2.5 Applicazioni ed estensioni di WordNet	33

3	Lo strumento utilizzato: MultiWordNet	35
3.1	La nascita di MultiWordNet	35
3.2	Struttura di MultiWordNet	36
3.3	Le relazioni alla base di MultiWordNet	37
3.3.1	Sinonimia, il concetto base dei synset	38
3.3.2	Antonimia	39
3.3.3	Iponimia/Iperonimia	40
3.3.4	Meronimia/Olonimia	42
3.3.5	Relazioni morfologiche	44
4	Esempi di utilizzo del MultiWordNet specialistico per il dominio della tutela delle acque	44
4.1	Ricerca dei termini	44
4.2	Visualizzazione delle relazioni tra termini	50
5	Conclusioni e sviluppi futuri	72
6	Ringraziamenti	73
	Riferimenti bibliografici	74

Elenco delle figure

1	Esempi di relazioni tra termini	18
2	Schermata presentata da MultiWordNet ad autenticazione effettuata .	45
3	Settaggio di browsing ed editing in lingua italiana del MultiWordNet per la tutela delle acque	46

4	Schermata presentata da MultiWordNet a seguito della selezione da parte dell'utente dell'etichetta "Search"	47
5	Scelta dei diversi criteri di ricerca in MultiWordNet	48
6	Risultato della ricerca del termine "acqua"	49
7	Risultato della ricerca del termine "acque sotterranee"	50
8	Visualizzazione degli iperonimi del termine "acque sotterranee"	52
9	Risultato della ricerca del termine "lago"	53
10	Risultato della ricerca degli iponimi del termine "lago"	54
11	Risultato della selezione del primo senso del termine "lago"	55
12	Menu a tendina per la visualizzazione dei termini in relazione col termine "lago"	56
13	Visualizzazione della lista di iponimi del termine "lago"	57
14	Visualizzazione del dettaglio del termine "lago polimittico", iponimo di "lago"	58
15	Visualizzazione della scheda relativa al termine "lago polimittico", iponimo di "lago"	59
16	Ricerca del termine "laghi naturali eutrofizzati"	60
17	Il termine "laghi naturali eutrofizzati" presenta un solo senso	61
18	Risultato della selezione del primo e unico senso del termine "laghi naturali eutrofizzati"	62
19	Visualizzazione della lista di iperonimi del termine "laghi naturali eutrofizzati"	63
20	Visualizzazione in MultiWordNet degli iponimi inseriti del termine "acqua"	65
21	Visualizzazione in MultiWordNet dei sinonimi inseriti del termine "acque costiere"	66

22	Visualizzazione in MultiWordNet dei sinonimi inseriti del termine "acqua potabile"	67
23	Ricerca del termine "acque superficiali"	68
24	Visualizzazione delle relazioni del termine "acque superficiali"	69
25	Visualizzazione degli antonimi inseriti del termine "acque superficiali"	70
26	Visualizzazione in MultiWordNet dell'antonimo inserito del termine "acque sotterranee"	71

Elenco delle tabelle

1	Lista dei venticinque synset alla base di WordNet	29
---	---	----

Abstract

Il presente lavoro di tesi ha avuto come obiettivo primario la realizzazione di un nuovo dizionario specialistico terminologico e semantico contenente i termini significativi all'interno del dominio delle normative sulla *tutela delle acque* e la sua implementazione informatica nel contesto di uno strumento software specializzato per la linguistica. Dopo aver completato l'implementazione del vocabolario sulla *tutela delle acque* in lingua italiana, si darà avvio alla sua implementazione nelle lingue inglese e spagnolo, con lo scopo di rendere fruibile il dizionario, anche in queste due importanti lingue comunitarie.

Introduzione

L'ISPRA (Istituto Superiore per la Protezione e la Ricerca Ambientale) opera da numerosi anni nel settore della *tutela delle acque* dall'inquinamento su scala nazionale ed in ambito comunitario; per *tutela delle acque* si intende l'insieme delle attività volte ad assicurare un elevato livello di qualità delle acque interne e delle acque marine italiane. Le attività in questione consistono sia nell'individuazione delle acque soggette ad inquinamento, che nel monitoraggio, nel risanamento delle acque inquinate e nel mantenimento delle acque pulite. La *tutela delle acque* viene effettuata attraverso l'adozione di strumenti normativi, nazionali e comunitari, che dispongono le azioni da svolgere e le modalità di esecuzione, l'assetto organizzativo e le rispettive competenze, la formulazione e la divulgazione dei risultati delle azioni svolte. In Italia la cultura della *tutela delle acque* è piuttosto recente: il Decreto Legislativo n. 152 del 1999 costituisce il primo approccio normativo integrale alla materia. Ancora più di recente si è provveduto ad aggiornare quella norma, rendendola più aderente a quanto prodotto sull'argomento in ambito comunitario, attraverso la promulgazione del Decreto Legislativo n. 152 del 2006, le cui disposizioni sono tuttora soggette a modifiche ed integrazioni. Stante lo scenario sopra delineato, viene a determinarsi l'opportunità di disporre di strumenti utili a supportare l'accrescimento ed il consolidamento della cultura della *tutela delle acque*. Per rispondere a questa esigenza, il servizio "Raccolta e Gestione Dati" del Dipartimento "Tutela delle Acque Interne e Marine" dell'ISPRA ha pensato di avviare un percorso di ricerca volto alla creazione di un dizionario terminologico della *tutela delle acque*, al fine di rendere disponibile una risorsa linguistica affidabile che potesse essere di supporto sia alle numerose istituzioni pubbliche centrali e territoriali competenti, sia al legislatore. L'obiettivo della presente tesi è stato, quindi, duplice, ed ha riguardato, in primo luogo, la realizzazio-

ne di un nuovo dizionario terminologico e semantico contenente i termini significativi all'interno del dominio delle normative sulla *tutela delle acque* e, in secondo luogo, l'implementazione informatica di questo dizionario specializzato. Per la costruzione del dizionario ci si è riferiti alle normative esistenti e riguardanti nello specifico la *tutela delle acque*; questa è stata una scelta di metodo, che sottolinea l'intento di mettere a disposizione dell'utenza una risorsa linguistica che avesse le caratteristiche della chiarezza e del rigore. Per quel che riguarda l'implementazione del dizionario terminologico sulla *tutela delle acque*, l'ISPRA ha avviato una collaborazione con la divisione TCC (Cognitive and Communication Technologies) della "Fondazione Bruno Kessler" di Trento, ente di ricerca della Provincia autonoma di Trento, che opera nel campo scientifico, tecnologico e delle scienze umane. La "Fondazione Bruno Kessler" ha fornito supporto formativo, tecnico e tecnologico per l'implementazione del dizionario terminologico della *tutela delle acque* attraverso l'utilizzo del database lessicale MultiWordNet (sviluppato e concesso in licenza dalla "Fondazione Bruno Kessler") [1].

Metodologia

Per la costruzione del dizionario terminologico della *tutela delle acque* ci si è riferiti alle normative esistenti e riguardanti nello specifico la *tutela delle acque*; il dizionario è stato costruito manualmente, derivando i termini in modo ragionato dalle normative. Per l'implementazione informatica del dizionario specialistico sulla *tutela delle acque* è stato utilizzato il database lessicale MultiWordNet, del quale si tratterà specificamente nei successivi capitoli. Allo stato attuale, parte del vocabolario sulla *tutela delle acque* è stato implementato in MultiWordNet; il lavoro di implementazione ha posto diverse questioni di natura semantica e linguistica, rendendo questo lavoro al contempo complicato e appassionante. In futuro si prevede di continuare con l'implementazione del vocabolario specialistico sulla *tutela delle acque* e di implementare anche le versioni inglese e spagnola del dizionario stesso.

1 Proposta per un dizionario terminologico specialistico sulla tutela delle acque

In questo capitolo si descriverà la realizzazione di un dizionario terminologico e semantico contenente i termini significativi all'interno del dominio delle normative sulla *tutela delle acque*. Due fondamentali criteri hanno fatto da guida alla creazione della terminologia della *tutela delle acque*: il primo è stato quello di assumere come riferimento i termini utilizzati nei testi legislativi attualmente in vigore; il secondo è stato quello di considerare testi legislativi anche nelle lingue inglese e spagnolo, per utilizzare appieno le potenzialità rese disponibili da MultiWordNet. Il primo criterio tende ad assicurare l'uso condiviso di termini propri, garantendo, per quanto possibile, l'assenza di ambiguità e di riferimenti e definizioni poco specialistici. Il secondo criterio consente di realizzare una terminologia utilizzabile negli ambiti internazionali, segnatamente quello comunitario e, attraverso l'adozione dello spagnolo, di avere un'apertura verso i grandi paesi extra-europei. Le attività di studio sperimentali avviate nel passato dall'ISPRA, consentono di stimare in diverse migliaia i termini specifici della terminologia della *tutela delle acque*, per la sola lingua Italiana. I termini sono stati tutti estratti dai testi legislativi senza ricorrere all'ausilio di strumenti automatici. Questa scelta discende, da un lato, dalla considerazione che i testi legislativi in materia costituiscono un corpus di dimensioni ridotte (400-500 pagine complessivamente); dall'altro, la specificità dei termini utilizzati è tale da richiedere comunque una dispendiosa attività di analisi di eventuali elenchi di termini prodotti dai tools per l'estrazione automatica delle parole. In sostanza, data la specificità dell'argomento *tutela delle acque* e della terminologia ad esso connessa, alla quale va ad aggiungersi la specificità del linguaggio normativo-legale, è stato indispensabile analizzare e vagliare attentamente e in modo manuale le normative. Si stima

in due anni il tempo necessario a realizzare la terminologia della *tutela delle acque* nelle tre lingue: italiano, inglese e spagnolo. Lo scopo prefisso infatti, è quello di realizzare il dizionario in italiano, e quindi individuare i termini della lingua inglese e spagnola semanticamente corrispondenti. L'inserimento della terminologia è stato e continuerà ad essere effettuato attraverso l'applicazione web MultiWordNet on-line; delle caratteristiche di questa applicazione si parlerà diffusamente nel terzo capitolo.

1.1 Le normative di riferimento sulla tutela delle acque

Per l'analisi e l'estrazione della terminologia sulla *tutela delle acque*, sono state considerate le seguenti normative:

- **Decreto Legislativo 30 maggio 2008 n.116:** dà attuazione alla direttiva 2006/7/CE, relativa alla gestione della qualità delle acque di balneazione, ed abroga la direttiva 76/160/CEE. Questo decreto è finalizzato a proteggere la salute umana dai rischi derivanti dalla scarsa qualità delle acque di balneazione anche attraverso la protezione ed il miglioramento ambientale ed integra le disposizioni di cui alla parte terza del decreto legislativo 3 aprile 2006, n. 152 e successive modificazioni. Il decreto inoltre, stabilisce disposizioni in materia di:
 - monitoraggio e classificazione della qualità delle acque di balneazione;
 - gestione della qualità delle acque di balneazione;
 - informazione al pubblico in merito alla qualità delle acque di balneazione.
- **Direttiva 91/676/CEE del Consiglio del 12 dicembre 1991:** direttiva del Consiglio Europeo relativa alla protezione delle acque dall'inquinamento provocato dai nitrati provenienti da fonti agricole. Questa direttiva si attua

in seguito al Decreto Legislativo del 21 Aprile 1999 sulle disposizioni di *tutela delle acque* dall'inquinamento. L'inquinamento causato dallo spandimento e dallo scarico di deiezioni del bestiame o dall'uso eccessivo di fertilizzanti è una delle realtà più preoccupanti di questi ultimi anni, poichè vanno ad interessare le falde acquifere per acqua potabile o di balneazione. Con questa direttiva e con opportune legislazioni si cerca di contenere questo problema senza andare a danneggiare un settore importantissimo quale l'agricoltura. Lo scopo di questa direttiva è di introdurre norme opportune per ridurre l'inquinamento delle acque causato direttamente o indirettamente dai nitrati di origine agricola e prevenire qualsiasi ulteriore inquinamento di questo tipo.

- **Decreto Ministeriale 16/06/2008 n. 131:** tale decreto, emanato dal Ministero dell'Ambiente e della Tutela del Territorio e del Mare, costituisce regolamento recante i criteri tecnici per la caratterizzazione dei corpi idrici (tipizzazione, individuazione dei corpi idrici, analisi delle pressioni) per la modifica delle norme tecniche del decreto legislativo 3 aprile 2006, n. 152, recante "Norme in materia ambientale", predisposto ai sensi dell'articolo 75, comma 4, dello stesso decreto.
- **Direttiva 91/271/CEE del Consiglio del 21 maggio 1991:** direttiva del Consiglio Europeo concernente il trattamento delle acque reflue urbane. La presente direttiva concerne la raccolta, il trattamento e lo scarico delle acque reflue urbane, nonchè il trattamento e lo scarico delle acque reflue originate da taluni settori industriali. Essa ha lo scopo di proteggere l'ambiente dalle ripercussioni negative provocate dagli scarichi di acque reflue.

- **Direttiva 2006/7/CE del Parlamento Europeo e del Consiglio del 15 febbraio 2006:** tale direttiva è relativa alla gestione della qualità delle acque di balneazione e abroga la direttiva 76/160/CEE.
- **Direttiva 2006/118/CE del Parlamento Europeo e del Consiglio del 12 dicembre 2006:** tale direttiva è relativa alla protezione delle acque sotterranee dall'inquinamento e dal deterioramento.
- **Decreto Legislativo 3 aprile 2006 n.152 "Norme in materia ambientale":** tale decreto legislativo ha come obiettivo primario la promozione dei livelli di qualità della vita umana, da realizzare attraverso la salvaguardia ed il miglioramento delle condizioni dell'ambiente e l'utilizzazione accorta e razionale delle risorse naturali. Questo decreto legislativo disciplina, in attuazione della legge 15 dicembre 2004, n. 308, le materie seguenti:
 - nella parte seconda, le procedure per la valutazione ambientale strategica (VAS), per la valutazione d'impatto ambientale (VIA) e per l'autorizzazione ambientale integrata (IPPC);
 - nella parte terza, la difesa del suolo e la lotta alla desertificazione, la *tutela delle acque* dall'inquinamento e la gestione delle risorse idriche;
 - nella parte quarta, la gestione dei rifiuti e la bonifica dei siti contaminati;
 - nella parte quinta, la tutela dell'aria e la riduzione delle emissioni in atmosfera;
 - nella parte sesta, la tutela risarcitoria contro i danni all'ambiente.
- **Decreto Legislativo 16 marzo 2009 n. 30 "Attuazione della direttiva 2006/118/CE, relativa alla protezione delle acque sotterranee**

dall'inquinamento e dal deterioramento”: tale decreto definisce misure specifiche per prevenire e controllare l'inquinamento ed il depauperamento delle acque sotterranee, quali:

- criteri per l'identificazione e la caratterizzazione dei corpi idrici sotterranei;
- standard di qualità per alcuni parametri e valori soglia per altri parametri necessari alla valutazione del buono stato chimico delle acque sotterranee;
- criteri per individuare e per invertire le tendenze significative e durature all'aumento dell'inquinamento e per determinare i punti di partenza per dette inversioni di tendenza;
- criteri per la classificazione dello stato quantitativo;
- modalità per la definizione dei programmi di monitoraggio quali-quantitativo.

1.2 Pianificazione delle attività

Il lavoro per l'estrazione, la classificazione e l'inserimento della terminologia sulla *tutela delle acque*, è stato suddiviso nelle seguenti fasi:

- Scelta dei domini di interesse
- Analisi delle normative ed estrazione dei termini
- Individuazione di un primo nucleo terminologico per l'inserimento in MultiWordNet
- Individuazione delle relazioni tra i termini del primo nucleo terminologico

- **Inserimento del primo nucleo terminologico e delle relazioni individuate in MultiWordnet**

Nei paragrafi successivi verranno illustrate nel dettaglio le fasi del lavoro sopra citate.

1.3 Scelta dei domini di interesse

In MultiWordNet il dominio è l'etichetta data all'insieme dei termini utilizzati nell'ambito di una specifica materia; normalmente molti termini hanno significati diversi se attribuiti a domini diversi. In generale, un termine può appartenere a più di un dominio, cioè assumere significato diverso all'interno di domini diversi. Per esempio, il termine "acque", presenta diversi significati: nel dominio dell'anatomia è sinonimo di "liquido amniotico", mentre nei domini *tutela delle acque*, *acqua*, *ambiente* il termine "acque" è sinonimo di "acqua" ed ha come definizione "le acque meteoriche e le acque superficiali e sotterranee". I domini individuati nell'ambito della competenza del Servizio "Raccolta e Gestione Dati" del Dipartimento "Tutela delle Acque Interne e Marine" sono tre: *tutela delle acque*, *acqua*, *ambiente*. MultiWordNet contiene al suo interno una lista di domini, alla quale è possibile aggiungere e togliere nuovi domini a seconda delle necessità. Nel caso del MultiWordNet per la *tutela delle acque*, è stato necessario aggiungere i sopra citati domini a quelli già esistenti.

1.4 Analisi delle normative ed estrazione dei termini

In questa prima fase si è lavorato all'analisi delle normative e all'estrazione da queste dei termini significativi relativi al dominio principale *tutela delle acque*, al quale sono stati aggiunti, per completezza e perchè largamente interessati dalla terminologia estratta dalle normative, i domini *acqua* e *ambiente*. In questa fase iniziale l'idea guida è stata l'individuazione e l'accorpamento di un congruo insieme di termini che

fossero correlati ai domini di interesse. Per costruire tale insieme, si è proceduto all'analisi delle normative, riguardanti la *tutela delle acque*, che vanno a regolamentare settori specifici di tale materia. In particolare sono state esaminate, in ordine cronologico, le normative citate nel capitolo precedente. Data la complessità e il grande quantitativo di termini contenuti, sono ancora in corso di analisi il Decreto legge 3 aprile 2006 n.152 e il Decreto Legge 16 marzo 2009 n.30. Al fine di realizzare un dizionario specializzato, nel quale compaiano soltanto termini specificamente correlati con i domini di interesse, sono stati applicati i seguenti criteri:

- Assegnare ad un termine la definizione specifica contenuta nella normativa, quando presente. Tale definizione specificherà la semantica del termine all'interno del dominio di riferimento. Nel caso in cui il termine non presenti una definizione all'interno di una normativa, verrà inserito in MultWordNet senza definizione.
- Includere nel dominio le denominazioni delle normative e di alcuni articoli.
- Poichè alcuni termini compaiono nelle normative al singolare ed al plurale, si è scelto di ridondare, estraendo lo stesso termine sia al singolare che al plurale.
- In prima istanza, si è scelto di segnalare comunque i termini non particolarmente significativi nell'ambito dei domini di interesse.
- Il dizionario non raccoglie solo termini singoli, ma anche frasi. È questo il caso della *collocazione*, che è una combinazione di più parole in cui il significato totale dipende dal significato delle singole parole, che acquistano significato proprio. Es: divieto permanente di balneazione. Nel caso specifico della normativa riguardante la *tutela delle acque* la presenza di collocazioni è piuttosto

consistente. Per ogni frase inserita, sono state inserite anche le singole parole della frase in stretta attinenza con i domini di interesse.

Per ogni normativa è stato prodotto un file excel contenente i termini estratti da tale normativa. Ogni file excel relativo ad una normativa è strutturato come una tabella con quattro campi: termine lessicale (singolo termine, collocazione o idioma), riferimento legislativo (relativamente alla normativa in cui compare il termine), definizione (solo se presente nella normativa dalla quale il termine è stato estratto), note (relazioni semantiche e lessicali individuate e informazioni/osservazioni ritenute utili a vario titolo). Di seguito vengono elencate le normative analizzate e i corrispondenti file excel prodotti:

- Decreto Legislativo 30 maggio 2008 n.116: Template_Termini1.xls
- Direttiva 91/676/CEE del Consiglio del 12 dicembre 1991: Template_Termini2.xls
- Decreto Ministeriale 16/06/2008 n. 131: Template_Termini3.xls
- Direttiva 91/271/CEE del Consiglio del 21 maggio 1991: Template_Termini4.xls
- Direttiva 2006/7/CE del Parlamento Europeo e del Consiglio del 15 febbraio 2006: Template_Termini5.xls
- Direttiva 2006/118/CE del Parlamento Europeo e del Consiglio del 12 dicembre 2006: Template_Termini6.xls
- Decreto Legislativo 3 aprile 2006 n.152: Template_Termini7.xls

1.5 Individuazione di un primo nucleo terminologico

In questa fase si è proceduto all'individuazione, tra tutti i termini estratti dalle normative, di un primo nucleo terminologico da inserire prioritariamente nel dizionario

terminologico realizzato tramite MultiWordNet; si è quindi lavorato all'identificazione dei *synset* (sta per *synonym set*, ossia insieme di termini sinonimi) da inserire in MultiWordnet. In MultiWordnet per *synset* si intende una lista di forme lessicali tra di loro in relazione di sinonimia. L'insieme dei termini appartenenti al primo nucleo terminologico è stato scelto tra tutti quelli rilevati con l'analisi terminologica, seguendo i seguenti criteri guida:

- **Ricorrenza:** sono stati considerati prioritariamente i termini più ricorrenti nelle diverse normative.
- **Presenza di definizione:** molti termini sono definiti in modo rigoroso all'interno delle normative nelle quali vengono citati ed utilizzati. La presenza di definizione del termine in una o più normative è stato un criterio discriminante per l'inserimento prioritario dei termini nel dizionario terminologico.

I termini individuati come primo nucleo di *synset* per l'immediato inserimento in MultiWordNet, sono circa 800.

1.6 Individuazione delle relazioni tra termini

In questa fase si è lavorato all'individuazione delle relazioni semantiche e lessicali presenti tra i termini scelti per l'inserimento prioritario nel dizionario. Le relazioni semantiche individuate sono le seguenti:

- **Iperonimia:** si ha iperonimia se il significato del secondo termine è incluso nel significato del primo. Es: "acque" è iperonimo di "acque di balneazione".
- **Iponimia:** si ha iperonimia se il significato del secondo termine include il significato del primo. Es: "corpo idrico naturale" è iponimo di "corpo idrico".

- **Meronymia:** la meronymia è un concetto di relazione semantica che indica un costituente "part of" (parte di), o un membro di qualcosa. Es: "affluente" è meronimo di "bacino idrografico".

Le relazioni lessicali individuate sono le seguenti:

- **Sinonimia:** la sinonimia indica la relazione che c'è tra due termini che hanno lo stesso significato all'interno dello stesso dominio di definizione. Es: nei nostri domini "acque potabili" è sinonimo di "acque destinate al consumo umano".
- **Antonimia:** l'antonimia indica la relazione che c'è tra due termini di significato opposto all'interno dello stesso dominio di definizione. Es: "acque superficiali" è antonimo di "acque sotterranee".

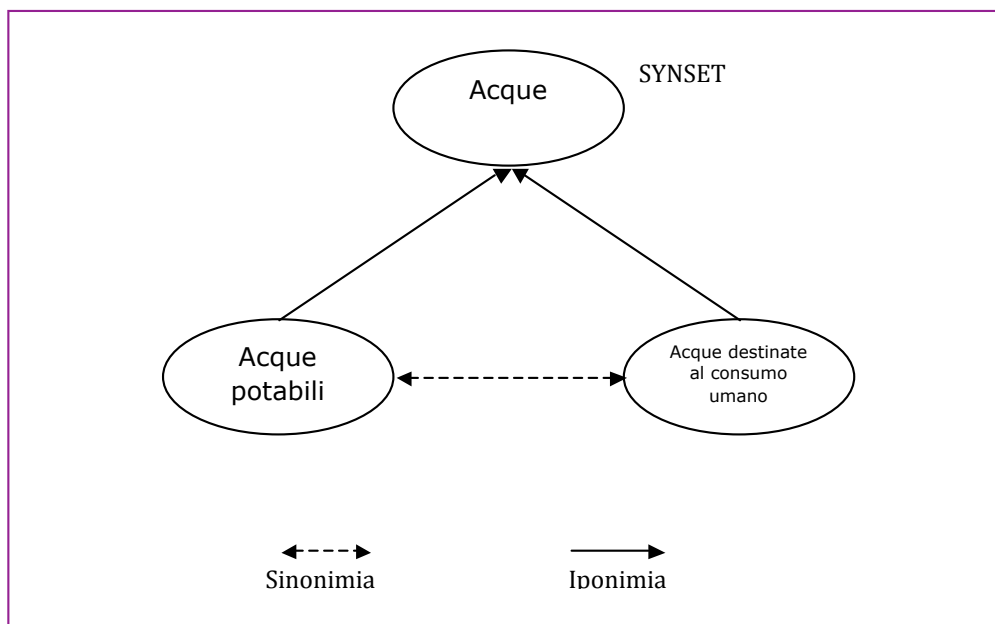


Figura 1: Esempi di relazioni tra termini

1.7 Inserimento dei termini

La prima esigenza ravvisata per l’inserimento dei termini in MultiWordNet è stata quella di aggiungere ai domini presenti al momento, quelli di: *tutela delle acque*, *ambiente*, *acqua*. Contestualmente all’inserimento dei *synset* sono state implementate le relazioni semantiche e lessicali precedentemente individuate; sono state inoltre inserite, per ogni *synset*, gli estremi delle normative di riferimento correlate, e, quando presenti, le definizioni. I *synset* inseriti fino ad oggi sono circa 400.

2 Il database lessicale WordNet

Nel presente capitolo verrà presentato in modo completo il database lessicale MultiWordnet; tramite questo potente strumento, che coniuga in se tutte le potenzialità di uno strumento automatico, implementando tutte le funzionalità indispensabili ad una trattazione linguistica rigorosa ed esaustiva è stato possibile implementare il dizionario specialistico sulla tutela delle acque. MultiWordNet, come detto precedentemente, è sviluppato e concesso in licenza per l’utilizzo, dalla “Fondazione Bruno Kessler di Trento”. Poichè MultiWordNet è nato come ulteriore sviluppo e specializzazione a partire dal database lessicale WordNet, e di WordNet conserva il nucleo fondante, nei paragrafi successivi verrà illustrato WordNet, in un percorso esplicativo che parte dalle ragioni per le quali questo strumento è nato, dalle necessità cui andava a dare una risposta, per arrivare alla descrizione della sua struttura, ai principi linguistici che la regolano e che guidano il suo utilizzo, facendo anche delle considerazioni sui suoi punti di forza e di debolezza.

2.1 Genesi di WordNet

WordNet nasce nel 1985 come risultato di un progetto al quale hanno partecipato linguisti e psicologi dell'Università di Princeton, sotto la guida di George Miller, uno dei più importanti psico-linguisti del '900. Dall'idea iniziale di fornire una ulteriore risorsa on line rispetto ad una semplice ricerca di tipo alfabetico, si è arrivati ad un vero e proprio dizionario basato sui principi della psicolinguistica [2]. All'inizio i ricercatori di Princeton pensarono di creare una ricerca concettuale nei dizionari che potesse costituire una risorsa in più rispetto alla ricerca alfabetica. Col procedere dei lavori le ambizioni divennero sempre più alte, fino a passare da un semplice meccanismo di ricerca, alla costituzione di un intero dizionario basato sui principi della psicolinguistica.

L'idea dei creatori di WordNet era di testare un'ipotesi secondo la quale il lessico è organizzato nella memoria umana, in aree separate a seconda della parte del discorso. I nomi, i verbi, gli aggettivi e gli avverbi, cioè le principali parti del discorso, sono strutturate nella memoria umana, secondo delle strutture indipendenti l'una dall'altra; ci saranno anche delle cose in comune, però ogni parte del discorso ha la sua strutturazione interna. L'altra ipotesi che si fece fu quella che un insieme di sinonimi fosse sufficiente per definire un concetto. E infatti in WordNet, ogni parte del discorso ha delle relazioni semantiche proprie distinte dalle altre; inizialmente WordNet non aveva scopi applicativi, era semplicemente un esperimento di psicolinguistica, con il tempo però crebbe talmente in dimensioni, che cominciò ad avere anche una rilevanza applicativa. Per esempio, la versione 1.6 di WordNet, che è una versione molto importante, perchè vari progetti si sono basati su di essa, conteneva già informazioni relative a 130.000 parole. Un dizionario di grosse dimensioni della lingua italiana contiene tra le 90.000 e le 100.000 parole. Grazie a questa grande

copertura e grazie al fatto che veniva distribuito gratuitamente senza vincoli commerciali o accademici, WordNet ebbe un enorme successo nel campo della linguistica computazionale. L'ultima versione di Wordnet è la 3.0; i vocaboli sono diventati 150.000 e i sensi delle parole riportati sono più di 200.000.

2.2 Forme di parola e significati

La semantica lessicale parte dal concetto chiave che una parola è un'associazione fra l'espressione della parola (forma e pronuncia) ed i concetti che la parola stessa esprime. La corrispondenza fra la forma di parola ed il concetto che esprime sono rappresentate nella matrice lessicale. Le relazioni sono del tipo "molti a molti" e i concetti di polisemia e sinonimia possono essere facilmente compresi se si pensa ai processi mentali: un lettore (o una persona che ascolta) deve scontrarsi con la polisemia ovvero deve scegliere, cioè quale significato associare alla forma di parola che ha recepito e, a sua volta, colui che scrive (o parla) deve decidere quale forma rende in maniera appropriata il significato che intende esprimere. Pariteticamente la matrice lessicale potrebbe essere rappresentata, in un diagramma, da due blocchi con due frecce che partendo da questi vanno in entrambe le direzioni. I due blocchi saranno chiamati rispettivamente "Word Meaning", cioè il significato della parola e "Word Form", cioè la forma della parola. Le due frecce indicano che colui che parla può partire da un significato "Word Meaning" e cercare una forma adeguata ad esprimerlo, oppure partendo da una "Word Form" ricercarne il giusto significato. Da quanto fino ad ora visto quindi, si può affermare che WordNet si è posto l'obiettivo di voler esprimere due diverse relazioni:

- relazioni semantiche, fra i significati;
- relazioni lessicali, fra le forme.

La costruzione della base di dati si è scontrata con l'esistenza di due teorie: la teoria costruttiva e la teoria differenziale. Secondo la teoria costruttiva un'accurata costruzione di un concetto deve essere supportata da un numero sufficiente di informazioni. Tali informazioni devono consentire di caratterizzarlo in modo da poterlo distinguere da altri possibili concetti lessicali e di fornirne una corretta definizione. La teoria differenziale, molto meno rigida, esprime il fatto che la rappresentazione di un concetto possa essere fatta solo con elementi che permettano di distinguerlo da altri. Per essere più chiari si può ricorrere ad un esempio: la parola *pianta*. Essa può avere i seguenti significati: nome generico che indica qualsiasi vegetale fornito di organi specializzati, proiezione orizzontale di un oggetto, parte inferiore del piede. Nella teoria costruttiva per differenziare i due significati dobbiamo fornire abbastanza informazioni in modo da distinguerli. In quella differenziale basta fornire una lista di forme che lo possano esprimere. Il significato *M* può essere espresso con una lista di forme (F_1, F_2, \dots) . In questo modo per ogni significato si ottiene una lista di forme fra di loro in relazione di sinonimia. L'insieme viene indicato appunto come *synonym set*, meglio conosciuto, nella sua forma contratta, come *synset*. Ritornando all'esempio precedente, per distinguere i due significati sarebbe stato sufficiente citarne due sinonimi: *vegetale* per il primo, e *mappa* per il secondo. Nel caso in cui non esista un sinonimo appropriato a differenziare quel significato da altri, si fa ricorso ad una *glossa*, ovvero una breve spiegazione del significato. Per il terzo significato dell'esempio proposto, si potrebbe utilizzare la seguente glossa: "parte inferiore del piede".

2.3 WordNet: organizzazione e principi di funzionamento

A una prima occhiata WordNet altro non è che un dizionario on-line, cioè un database lessicale che può essere letto tramite un computer [3]. Il primo vantaggio che

deriva dall'utilizzo di un calcolatore è una velocità di ricerca superiore a quella di un essere umano a contatto con un dizionario su supporto cartaceo. Tuttavia risulta evidente che utilizzare un elaboratore come semplice "cercatore di informazione" è non sfruttare a pieno le potenzialità a disposizione, risulta alquanto limitativo. Per questa ragione, realizzando WordNet, si è ottenuto molto di più di un semplice dizionario on-line: con WordNet è stato pensato e costruito un modo completamente nuovo di strutturare le informazioni. WordNet non è organizzato alfabeticamente, ma su base concettuale; le informazioni sono memorizzate in base al significato e collegate fra loro per mezzo di una serie di riferimenti che permettono di soddisfare ad una ricerca fornendo non solo la risposta cercata, ma anche informazioni aggiuntive sia a livello generale che a livello specifico, dettagliato[4]. WordNet costituisce un "navigatore" all'interno dell'informazione: esso infatti aiuta l'utente nel reperimento delle informazioni che gli abbisognano, guidandolo tanto nella risalita verso la generalità quanto nella discesa verso la particolarizzazione, verso lo specifico. Alla base di tutto questo vi è un sistema per memorizzare l'informazione che passa da uno schema classico del tipo nome, definizione ad uno più articolato nome, definizione, puntatori dove i puntatori sono riferimenti ad altre informazioni contenute nel database. WordNet inoltre, presenta delle evidenti differenze rispetto ai comuni dizionari. La prima è la divisione del lessico in cinque categorie:

- **nomi;**
- **verbi;**
- **aggettivi;**
- **avverbi;**
- **parole funzione.**

Questa suddivisione è frutto di una diversa organizzazione del lessico a seconda della categoria in cui ci si trova. La struttura con cui i nomi vengono memorizzati è differente da quella con cui si memorizzano gli aggettivi o i verbi. Questa situazione è dovuta alle differenti relazioni con cui gli esseri umani sono portati a collegare i nomi (iperonimia/iponimia) e gli aggettivi (similitudine/antonimia). Il vantaggio di questa suddivisione consiste quindi nella presenza di un'organizzazione chiara e consistente all'interno delle diverse categorie. Di contro è presente uno svantaggio: la presenza di una certa ridondanza. L'obiettivo di WordNet e della presente tesi è di organizzare le informazioni lessicali secondo il significato e non secondo la forma e cercare di farlo in maniera corretta e non ridondante. Per capire come sia possibile far questo è necessario capire la struttura base su cui si fonda WordNet. L'elemento di base, quello che può essere definito il mattone costitutivo di WordNet, è la parola.

Definizione 1: La parola è un'associazione fra un concetto lessicale e la stringa che lo rappresenta.

La domanda che ci si pone dopo aver definito cos'è una parola è la seguente: qual'è la natura e l'organizzazione dei concetti lessicali che le parole esprimono? La risposta più immediata è che tale organizzazione dovrà riflettere la presenza di due elementi in una parola: il concetto lessicale o significato ("word meaning") e la stringa o forma ("word form"). Una prima osservazione porta a stabilire che la relazione fra le forme di parola e i significati è del tipo "molti a molti"; infatti ad una stringa possono corrispondere più significati (polisemia) e ad un significato possono essere associate più forme (sinonimia). Risulta evidente come polisemia e sinonimia siano aspetti complementari della stessa mappa e, come si vedrà in seguito, essi risultano fondamentali per l'organizzazione delle parole in WordNet.

Si può fare un esempio utilizzando WordNet 2.0. Se si effettua la ricerca della parola "dog" si ottengono sette possibili significati. A questo punto si può navigare nell'informazione in vari modi, per esempio spostandosi verso la generalità alla ricerca degli iperonimi e ottenendo la seguente gerarchia:

dog

⇒ *canine*

⇒ *carnivore*

⇒ *placental, plac.mammal, eutherian, euth.mammal*

⇒ *mammal*

⇒ *vertebrate, craniate*

Dalle precedenti considerazioni si intuisce come sia necessario rappresentare in WordNet due diverse relazioni:

- relazioni semantiche fra i significati;
- relazioni lessicali fra le forme.

Questi due concetti vengono utilizzati per memorizzare le parole e collegarle fra loro. Il metodo con cui ci avviene dipende dalla teoria scelta per l'organizzazione della base di dati. Esistono infatti due teorie: costruttiva e differenziale. Secondo la prima teoria la rappresentazione di un concetto deve presentare elementi a sufficienza per la costruzione del concetto stesso. Per identificare e caratterizzare un significato deve essere portato un numero di informazioni sufficiente, non solo a distinguerlo da altri possibili concetti lessicali, ma a definirlo in maniera corretta. Di seguito

viene presentato un esempio, considerando una parola che abbia più di un significato (polisemia) della parola "parco". Tale forma può assumere due significati:

1. terreno boscoso e piuttosto esteso, spesso recintato ed adibito a usi particolari;
2. sobrio, frugale, parsimonioso.

Se definendo il parco (riferito al primo significato) si specificasse che può essere recintato, si potrebbe differenziarlo senz'altro dal secondo, in quanto la parola recintato non ha alcun senso riferita al significato sobrio, ma non potremmo risalire al senso perchè un campo o un cortile può essere recintato senza essere un parco. In una definizione costruttiva si devono fornire abbastanza informazioni da poter costruire il significato del termine esaminato. Invece la seconda teoria, più semplice, si preoccupa unicamente della presenza di conoscenze che permettano di distinguere due elementi differenti. In questo senso viene chiamata differenziale. L'idea è quella di separare i vari significati (ovviamente la differenza delle forme comporta un immediata distinzione fra esse). In pratica si tratta di rappresentare un significato M_x per mezzo dell'insieme delle forme che lo supportano F_1, F_2, \dots, F_n . Così facendo si ottiene, per ogni significato un insieme di forme di parola in relazione di sinonimia (hanno tutte lo stesso significato); tale insieme è detto synset dalla contrazione di synonym set (insieme di sinonimi) ed è rappresentato dalla stringa x_0, x_1, \dots, x_n dove x_0, \dots, x_n sono gli elementi dell'insieme. Da sottolineare che i synset non spiegano cosa sono i concetti, ma affermano che i concetti esistono; per spiegare esattamente un concetto, deve essere presente anche una sua definizione. WordNet è una base di dati che si poggia, come sarà più chiaro in seguito, sull'associazione fra concetti, ma queste associazioni vengono fatte studiando le relazioni lessicali e semantiche esistenti. Le relazioni fondamentali sulle quali si poggia WordNet, sono le seguenti:

- **Sinonimia;**
- **Iponimia/Iperonimia;**
- **Meronimia/Olonimia;**
- **Antonimia;**
- **Relazioni morfologiche.**

Queste stesse relazioni sono alla base anche del database MultiWordNet, che da WordNet deriva e che è stato utilizzato in questa tesi per l'implementazione del vocabolario specialistico sulla *tutela delle acque*; di MultiWordNet e delle relazioni lessicali e semantiche di cui sopra, si parlerà diffusamente nel capitolo successivo, quando si tratterà di MultiWordNet. Nei paragrafi successivi invece, si tratteranno le categorie lessicali di WordNet.

2.4 WordNet e le categorie lessicali

Discutendo i principi alla base della costruzione di WordNet è stato evidenziato come sia necessario, in diverse situazioni, fare una separazione fra le diverse categorie lessicali. Alla base di questa scelta ci sono varie motivazioni, che possono essere riassunte con le seguenti considerazioni:

- la rappresentazione dei concetti avviene per mezzo dei synset, cioè insiemi di sinonimi, ma la sinonimia è una relazione che non può essere definita trasversalmente alle diverse categorie;
- all'interno delle diverse categorie lessicali è presente un'organizzazione gerarchica distinta, basata su relazioni base diverse a seconda della categoria considerata.

Di seguito si spiegherà come procedere all'organizzazione dei synset all'interno di ogni categoria, con particolare interesse per i nomi.

2.4.1 I nomi

I nomi sono organizzati secondo una struttura gerarchica con ereditarietà in base alla relazione di iponimia/iperonimia. Tale struttura ad albero permette di eliminare il rischio delle cosiddette definizioni circolari, come già si è spiegato in precedenza. La relazione di iponimia/iperonimia è quella che caratterizza la formazione dell'albero e, perciò, si avrà un maggior livello di generalità risalendo l'albero, mentre scendendo ai livelli più bassi dell'albero si aumenta il livello di specificità. Si è soliti distinguere un livello base, al di sopra del quale le definizioni sono generali, mentre al di sotto di questo livello, le definizioni diventano particolareggiate. Ogni synset è collegato con tramite puntatori sia ai suoi iponimi che ai suoi iperonimi. Gerarchie di questo tipo sono molto utilizzate in database di grosse dimensioni, presentando il vantaggio di risparmiare spazio evitando le ridondanze. Un modo di costruire l'albero è considerare tutti i nomi appartenenti ad un'unica gerarchia. Per far questo è necessario porre come radice qualcosa di molto astratto e partire da questo punto. All'atto pratico questo modo di procedere genera numerosi concetti, ognuno dei quali porta con sé pochissima informazione semantica. Per ovviare a questo si è attuata una scelta alternativa: partizionare l'insieme dei nomi per mezzo di un insieme di synset primitivi, ognuno dei quali radice unica di una gerarchia indipendente; questi synset corrispondono a campi semantici distinti. Un altro vantaggio conseguente alla loro creazione è la minor dimensione dei singoli file lessicografici, cioè dei file contenenti i synset. Le venticinque radici alla base dei nomi in WordNet sono riportate in Tabella 1. Non è facile esplicitare dei synset di partenza con la proprietà che ogni parola del vocabolario trovi posto in una gerarchia discendente da una di queste radici.

art, action, activity	natural object
animal, fauna	natural phenomenon
artifact	person, human being
attribute, property	plant, flora
body, corpus	possession
cognition, knowledge	process
communication	quantity, amount
event, happening	relation
feeling, emotion	shape
food	state, condition
group, collection	substance
location, place	time
motive	

Tabella 1: Lista dei venticinque synset alla base di WordNet

Le soluzioni possono essere molteplici ed è difficile poter affermare di aver risolto il problema in modo ottimo. Per questo è giusto dire che i synset che rappresentano le venticinque radici sono stati eletti a quel ruolo per scelta, in modo da creare dei nodi di partenza da cui le gerarchie potessero essere sviluppate. È doveroso sottolineare che tali gerarchie non sono mutuamente esclusive; vale a dire che sono presenti numerosi riferimenti incrociati fra significati presenti in diverse gerarchie. Ad esempio si consideri la parola "man" (uomo): essa appartiene alla gerarchia che ha come radice person, human being, ma costituisce pure un iponimo del termine "essere vivente" che è presente nella gerarchia originata da natural object. In questo caso è presente una relazione fra synset appartenenti a differenti gerarchie confermando che esse non sono mutuamente esclusive. Vista la struttura gerarchica governata dalla relazione di iponimia/iperonimia, le differenze fra i vari significati sono formalizzate esplicitando le caratteristiche che distinguono un concetto da un altro. Tali caratteristiche distintive, che vanno trattate separatamente, sono di tre tipologie:

- **Attributi;**
- **Parti (meronimia);**
- **Funzioni.**

È importante notare che, di queste tre, solo la caratteristica denominata "parti" è rappresentata da altri nomi, mentre gli "attributi" sono aggettivi e le "funzioni" sono verbi. Questo fa dedurre che solo la meronimia può essere descritta da puntatori interni alla categoria nomi, mentre le altre due caratteristiche possono essere esplicitate solo con puntatori verso categorie distinte. Questo non è stato realizzato nelle prime versioni di WordNet però è giusto tenerne conto poichè fornisce potenzialità molto elevate. Per quanto riguarda la meronimia, essa è rappresentata senza difficoltà grazie a puntatori verso altri nomi. È necessario evidenziare che vi sono tre tipi di puntatori, in quanto vengono prese in considerazione tre diversi tipi di relazioni. Un'ultima relazione non necessaria all'organizzazione dei nomi, ma che può essere utile, è l'antonimia. Essa viene inclusa in WordNet rimanendo opzionale. La rete di nomi che si ottiene considerando le tre relazioni semantiche definite (iperonimia, meronimia, antonimia) è una rete piuttosto complessa.

2.4.2 Gli aggettivi

In WordNet vengono distinti aggettivi di tre diverse tipologie:

- **Aggettivi descrittivi;**
- **Aggettivi relazionali;**
- **Aggettivi "reference-modifying".**

Un aggettivo descrittivo associa un valore a un attributo del nome cui esso è associato. Si pensi per esempio a "casa grande" : "grande" è un aggettivo descrittivo riferito all'attributo "dimensione" del nome casa. Questa tipologia di aggettivi non è organizzata secondo un albero o una gerarchia come per i nomi, ma per mezzo di uno spazio dimensionale di cui sono noti gli estremi. In riferimento all'attributo "grandezza" viene definita una scala di valori (spazio unidimensionale) estesa, ad esempio, fra gli estremi "infinitesimo" ed "enorme". La relazione fondamentale per la definizione di queste scale è l'antonimia. Gli aggettivi relazionali sono aggettivi che derivano dai nomi (ad esempio atomico deriva da atomo). Essi non si riferiscono ad una proprietà del nome che modificano e non sono quindi associati ad un attributo. Aggettivo relazionale significa qualcosa del tipo "relativo/pertinente a, associato con" un nome. Essi inoltre non sono organizzati come gli aggettivi descrittivi perchè per essi non è significativa la relazione di antonimia. Questi aggettivi sono mantenuti in una lista distinta e puntano al nome da cui derivano (puntatori trasversali alle categorie). Per quanto riguarda gli aggettivi "reference-modifying", si può dire che un numero piuttosto piccolo di aggettivi, comprendente former (precedente) e al leged (fatto supposto vero senza che vi sia una dimostrazione di esso), costituiscono una classe chiusa che viene mantenuta separata dagli altri aggettivi. Spieghiamo che tipologia di aggettivi costituiscono con un esempio, consideriamo la frase seguente: "Ieri ho incontrato un mio vecchio amico". L'aggettivo vecchio può avere due collocazioni: essere riferito alla persona amica e indicare che tale persona è anziana ("referent-modifying", modifica il referente); essere riferito all'amicizia indicando che è di lunga data ("reference-modifying", modifica la referenza). In quest'ultimo caso, la tipologia di aggettivi in esame, cioè quelle parole che modificano la referenza che passa dall'amico all'amicizia. Di seguito un altro esempio per chiarire meglio il concetto: se si dice "il preside precedente" si modifica la persona cui ci si riferisce,

passando dal preside a colui che lo precedeva. Come ultima osservazione si può aggiungere che spesso questi aggettivi funzionano come avverbi (ad esempio: "il mio insegnante precedente" e "lui precedentemente fu il mio insegnante"), sono sempre attributivi e mai predicativi e spesso hanno antonimi diretti.

2.4.3 I verbi

I verbi, compresi i "phrasal verbs" (frasi verbali come, ad esempio, stare per, accingersi a...), hanno in WordNet un'organizzazione gerarchica. Essi sono divisi in due categorie:

- verbi che denotano azioni o eventi;
- verbi che denotano stati.

I primi sono a loro volta suddivisi in quattordici tipologie corrispondenti a quattordici diversi gruppi semantici. Per quanto riguarda i secondi, essi costituiscono un raggruppamento unico. Le relazioni in base a cui sono organizzati i verbi sono due:

- **implicazione;**
- **opposizione.**

Vengono considerate quattro diverse implicazioni, in dipendenza dal rapporto temporale fra il verbo implicante e il verbo implicato; in particolare se c'è inclusione temporale fra i due, si parla di co-occorrenza o troponimia (ad esempio: zoppicare implica camminare) e questa relazione corrisponde a quella che per i nomi è detta iponimia. In pratica si dice che un verbo è troponimo di un altro se si può dire che il primo verbo è un modo dell'altro. Questa relazione è particolarmente importante perchè permette la costituzione di una struttura gerarchica che è tuttavia diversa

da quella per i nomi in quanto all'interno di un campo semantico non tutti i verbi possono essere riuniti sotto un unico nodo. Per quanto riguarda l'opposizione va evidenziato che viene codificata solamente l'opposizione lessicale, non quella semantica. Per quanto riguarda l'informazione associata ad ogni verbo in WordNet è possibile dire che ne viene specificata la struttura predicato-argomento: sono attribuiti ruoli tematici ai nomi che fungono da argomenti del verbo e vengono specificate le proprietà semantiche riguardanti le classi di nomi che possono costituire l'argomento per un verbo. Infine, per ogni verbo del synset, vengono descritte una o più strutture (frames) che specificano le caratteristiche di sottocategorizzazione dei verbi (indicando le frasi in cui possono comparire).

2.4.4 Gli avverbi

Gli avverbi sono memorizzati in maniera molto semplice in quanto non è presente né una struttura gerarchica (come per i nomi e per i verbi) né un'organizzazione in gruppi (come per gli aggettivi). Per ogni avverbio sono indicati i sinonimi e gli antonimi e, quando derivano da un aggettivo, questo viene specificato. Gli avverbi costituiscono una categoria molto eterogenea (anche per la presenza di numerosissime locuzioni avverbiali), ma non è mai stato fatto alcun tentativo allo scopo di suddividerli in gruppi distinti.

2.5 Applicazioni ed estensioni di WordNet

Nel tempo WordNet è diventato il tool ideale in molti campi di ricerca, grazie a diverse sue qualità: l'essere open source, quindi liberamente fruibile in modo del tutto gratuito, l'essere dotato di una documentazione esauriente e l'avere grande potenzialità nell'ambito del NLP (Natural Language Processing). Si può quindi dividere il lavoro di sviluppo e ricerca su WordNet in due grandi rami: da una parte

l'implementazione e l'arricchimento dello strumento stesso, dall'altra il suo utilizzo da parte dei sistemi NLP. Le applicazioni che utilizzano WordNet nell'ambito di NLP sono le seguenti:

- **Information Retrival e Extraction:** queste operazioni di recupero ed estrazione delle informazioni da una sorgente, in genere la rete Internet, sono strettamente legate all'organizzazione e rappresentazione del sapere nella rete. Una delle aree della ricerca attualmente più attive è l'applicazione degli strumenti di intelligenza artificiale alla ricerca delle informazioni in rete. In questo caso, WordNet viene usato come strumento di conoscenza linguistica atto a rappresentare ed interpretare il significato dell'informazione.
- **Disambiguation:** la risoluzione di ambiguità semantiche dei termini consiste nel determinare in maniera automatica il significato più appropriato di una parola in base al contesto nel quale essa si trova. WordNet, in quanto rete semantica, viene utilizzato come strumento per determinare il senso di una parola sulla base delle relazioni semantiche con altre parole.
- **Distanza Semantica:** il concetto di distanza semantica è direttamente connesso a quello di similarità semantica, che fornisce appunto una misura della similarità semantica fra due parole. Si può affermare che il calcolo della similarità semantica fra due lemmi si basa sulla lunghezza del cammino necessario a percorrere la distanza che li separa dal concetto minimo comune.

Nell'ambito dei lavori di arricchimento di WordNet, il più importante per l'ambito di ricerca di questa tesi, riguarda la creazione di database specifici ove ci si concentra su un certo dominio, costruendo gerarchie fra concetti ad alta specializzazione, con un uso limitato di relazioni linguistiche e lessicali. A partire da WordNet, sono stati

implementati numerosi database lessicali per la lingua Italiana: Economic WordNet per il campo dell'economia, Jur-WordNet per il campo giuridico, ArchiWordNet per l'Architettura e il Maritime-WordNet per termini legati alla navigazione e al commercio marittimo. Il presente lavoro di tesi rappresenta la descrizione del lavoro di ricerca svolto proprio nella direzione di arricchimento e specializzazione di WordNet, per la creazione di un database lessicale sulla *tutela delle acque*.

3 Lo strumento utilizzato: MultiWordNet

Come visto nel precedente capitolo, MultiWordNet è un'estensione di WordNet che ne rappresenta l'apertura al multilinguismo; per l'implementazione del dizionario specialistico sulla tutela delle acque, è stato utilizzato questo strumento, che ha consentito una rappresentazione agile e completa delle informazioni. Nei paragrafi successivi verrà presentato il database lessicale MultiWordNet, se ne descriveranno le basi teorico-pratiche, la struttura, i principi di funzionamento e si forniranno diversi esempi di utilizzo.

3.1 La nascita di MultiWordNet

Una volta realizzato, WordNet ha dimostrato di essere un supporto valido per la ricerca di informazioni e non un semplice dizionario informatico. Per le potenzialità espresse e per le applicazioni, anche commerciali, che questa base di dati ha dimostrato di poter soddisfare, si è pensato ad un progetto più ambizioso: un WordNet multilingue e multitematico. L'obiettivo è diventato costruire un sistema governato dai principi di WordNet che riunisse la conoscenza delle diverse lingue e dei diversi ambiti tematici. Da questa idea è nato MultiWordNet. MultiWordNet è un progetto attualmente in fase di sviluppo da parte della divisione TCC (Cognitive and Commu-

nication Technologies) della "Fondazione Bruno Kessler" di Trento. Esso prevede la realizzazione di un lessico generico su larga scala per l'italiano basato sulla versione inglese di WordNet. Più propriamente l'obiettivo del progetto MultiWordNet è creare un oggetto multilingue in cui siano considerate tanto le relazioni intralinguistiche interne alla rappresentazione delle diverse lingue, quanto le relazioni interlinguistiche che costituiscono il collegamento fra lingue diverse. Per rendere possibile questo è necessario ipotizzare che le strutture concettuali di livello lessicale presenti nelle diverse lingue siano confrontabili e, almeno in parte, sovrapponibili.

3.2 Struttura di MultiWordNet

A livello di struttura MultiWordNet è organizzato come un insieme di reti monolingue messe assieme per mezzo di una gerarchia comune (denominata WN-comune). Questa costituisce lo scheletro concettuale su cui sono attaccati i diversi livelli corrispondenti alle varie lingue. WN-comune ricalca, nella sostanza, WordNet inglese tranne che in alcune peculiarità. I livelli monolingue hanno una certa indipendenza, che è totale nella rappresentazione delle relazioni lessicali della lingua considerata. Per quanto riguarda le relazioni semantiche esse vengono ereditate dalla gerarchia comune se compatibili, altrimenti vengono espresse autonomamente, mantenendo comunque una corrispondenza fra i nodi della gerarchia monolingue e i nodi di WN-comune. I livelli monolingue sono costituiti da due parti:

- La parte corrispondente alla gerarchia WN-comune;
- La parte differente dalla gerarchia WN-comune.

La parte corrispondente alla gerarchia WN-comune contiene solo *synset* e puntatori alla gerarchia comune. La parte differente dalla gerarchia WN-comune contiene un numero maggiore di informazioni perchè è indipendente dalla gerarchia comune

(questa parte viene anche detta WN-differenza). Il mattone fondamentale di MultiWordNet è il *synset*; (*synset* sta per *synonym set*, cioè insieme di sinonimi) e l'idea è che questo insieme di sinonimi esprime un concetto. Più parole hanno lo stesso significato e rappresentano un significato. I componenti di questi insiemi di sinonimi sono le *lexical unit*, in italiano *unità lessicali*. Una *lexical unit* è o una parola o un'espressione che si considera come una unità concettuale: quindi, per fare un esempio, una *lexical unit* è "casa", ma anche "casa di cura", laddove quello che noi intendiamo con "casa di cura" è un termine formato da più parole che esprime però un concetto unitario. In MultiWordNet, per termine o parola si intende sempre una *lexical unit*. Una volta costruiti questi blocchi, queste unità concettuali, MultiWordNet li mette in relazione tra loro, cioè definisce esplicitamente delle relazioni che sono di due nature:

- **Relazioni lessicali:** le relazioni che intercorrono tra parole (la sinonimia è una relazione lessicale);
- **Relazioni semantiche:** le relazioni che intercorrono tra *synset* (l'antonimia è una relazione semantica).

3.3 Le relazioni alla base di MultiWordNet

MultiWordNet si basa sulle relazioni semantiche fra concetti, fra le quali la sinonimia gioca un ruolo fondamentale, ma non è l'unica ad essere utilizzata. Le relazioni che verranno descritte in seguito saranno quelle fondamentali, anche se non le uniche implementate da MultiWordNet:

- **Sinonimia**
- **Antonimia**

- Iponimia/Iperonimia
- Meronimia/Olonimia
- Relazioni Morfologiche

Nei prossimi paragrafi verranno illustrate nel dettaglio le relazioni sopra citate.

3.3.1 Sinonimia, il concetto base dei synset

La prima relazione importante al fine dell'organizzazione di MultiWordNet è la relazione di sinonimia. Questa relazione permette, non tanto di legare diversi significati, ma di rappresentarli all'interno del dizionario. Ogni concetto è identificato da un *synset*, che è proprio l'insieme costruito unendo elementi fra loro sinonimi. Di seguito vengono fornite delle definizioni di sinonimi

Definizione 2: due concetti sono fra loro sinonimi se la sostituzione di uno con l'altro non cambia il valore di verità della frase nella quale viene fatta la sostituzione.

Questa è una definizione molto stringente e, in base ad essa, ben pochi termini si possono considerare sinonimi. Se la tenessimo presente per la costruzione del nostro dizionario non si riuscirebbe ad ottenere *synset* formati da una o due forme di parola. Per questo si è scelto, per la realizzazione di MultiWordNet, di adottare un concetto di sinonimia più adatto allo scopo, ridefinendo la sinonimia come segue:

Definizione 3: due concetti sono fra loro sinonimi in un contesto linguistico C se la sostituzione di un concetto con l'altro nel contesto C non ne altera il valore di verità.

Oltre a questa definizione in termini di vero/falso, esiste un'altra scuola di pensiero, secondo la quale i sinonimi possono essere pensati anche secondo un concetto di similarità. Così una relazione di similarità semantica è sufficiente a rendere due concetti fra loro sinonimi. In molti casi infatti, non si può considerare la sinonimia come un fatto discreto, vero o falso, come le definizioni sopra scritte affermano. Quello che si vuole è ottenere degli insiemi di forme di parole che permettano di distinguere un significato da un altro e, per raggiungere questo scopo, la relazione semantica di similarità risulta sufficiente. Così la sostituzione della parola *pianta* con *mappa* non ne altera il significato in topografia, ma in altri contesti la sostituzione di una parola con l'altra potrebbe essere del tutto inappropriata. Per esempio: "albero", "arbusto" e "pianta" non possono essere fra loro sostituite, in quanto esistono anche piante che non sono necessariamente alberi o arbusti, ma, secondo una definizione di similarità, essi appartengono allo stesso synset, in quanto consentono la distinzione da altri significati. Nella costruzione del MultiWordNet sulla *tutela delle acque* si è scelto di seguire la Definizione 3: nel contesto linguistico normativo di riferimento del dizionario sulla *tutela delle acque* un concetto risulta sinonimo di un altro, se la sostituzione di un concetto con l'altro non altera il valore di verità.

3.3.2 Antonimia

Una relazione alla quale non sempre è facile dare una definizione, ma che risulta molto familiare, è l'antonimia. L'antonimo di una parola x viene definito quasi sempre come $notx$. Così "ricco" e "povero" sono fra loro antonimi, anche se essere non ricchi non implica necessariamente essere poveri: infatti, si può essere né ricchi, né poveri. Le due parole "ricco" e "povero" sono antonime, ma la negazione di "ricco", cioè "non ricco" non è un antonimo della parola di partenza. Per questo la relazione di antonimia non è una semplice relazione simmetrica come sembrerebbe. Per renderla

consistente, essa viene considerata come una relazione sulle forme di parola e non sui significati, quindi risultano antonime solo le forme di parola che lo sono per costruzione. La relazione di antonimia non è fondamentale nel sistema di relazioni rappresentate in MultiWordNet, ma viene comunque esplicitata quando presente.

3.3.3 Iponimia/Iperonimia

La relazione di iponimia/iperonimia è una delle più importanti presenti in MultiWordNet, costituisce la base della strutturazione dei nomi in MultiWordNet ed è detta anche relazione di specializzazione/generalizzazione. Le relazioni di iponimia e iperonimia sono relazioni fra significati secondo la definizione seguente:

Definizione 4: un concetto rappresentato dal synset x_1, x_2, x_3 , viene detto iponimo del concetto rappresentato dal synset y_1, y_2 , se si può accettare una frase costruita come: "un x è un (un tipo di) y ".

La relazione di iponimia/iperonimia è transitiva ed antisimmetrica. La struttura derivante dall'applicazione della relazione di iponimia/iperonimia ai synset è gerarchica in quanto da un concetto padre derivano più concetti figli che condividono delle caratteristiche base con il padre, ma presentano delle peculiarità proprie che li rendono particolari. La relazione di iponimia/iperonimia infatti, genera una struttura semantica gerarchica secondo la quale gli iponimi (concetti figli) stanno sotto il proprio iperonimo (concetto padre). Secondo questo criterio, la definizione comune di un nome è in genere data attraverso un termine più generale che lo descrive (iperonimo) e dall'elenco di caratteristiche che lo distinguono da altri. La relazione di iponimia/iperonimia introduce quindi il concetto di ereditarietà; un figlio eredita dal padre tutte le caratteristiche, aggiungendone poi altre che permettono di clas-

sificarlo rispetto al padre e che lo distinguono dagli altri figli. Così ad esempio, un ciliegio è un albero e si distingue dagli altri alberi per la durezza del legno, il fatto che è un albero da frutto, la forma delle foglie, il tipo di radici, ecc... Si è detto che dall'applicazione della relazione di iponimia/iperonimia ai *synset* viene generata una struttura gerarchica; questa gerarchia risulta molto utile per la costruzione di una base di dati, per due motivi principali:

- permette di ridurre la ridondanza dei dati, in quanto le informazioni e le caratteristiche condivise dal padre con tutti i suoi figli, vengono immagazzinate una sola volta solo per il padre;
- agevola la ricerca, in quanto permette di passare facilmente sia dalla generalità alla specificità, scendendo le gerarchie, sia dalla specificità alla generalità, risalendo le gerarchie.

La struttura gerarchica secondo la quale i nomi vengono organizzati inoltre, è ad albero, e questo permette di eliminare il rischio delle cosiddette definizioni circolari (esempio: la parola W_a è definita in base alla parola W_b e la parola W_b è definita in base alla parola W_a). Un esempio di iperonimia (detta anche relazione "IS A") è il seguente: "esercizio commerciale" è iperonimo di "negozio", ma anche di "banca-rella" e di "chiosco"; quindi "negozio", "banca-rella" e "chiosco" sono delle tipologie di "esercizio commerciale", ossia delle sue specializzazioni. Ogni iponimo di "esercizio commerciale" eredita le caratteristiche che contraddistinguono il padre, ma presenta anche delle caratteristiche strettamente proprie, che lo differenziano dagli altri figli dello stesso padre.

3.3.4 Meronimia/Olonimia

La relazione di meronimia/olonimia è una relazione semantica, ed è chiamata anche relazione "parte di" o "parte/tutto".

Definizione 5: Si dice che un synset x_1, x_2 , è meronimo di un synset y_1, y_2 , se si può rispondere affermativamente alla domanda ' x_1, x_2 , è una parte di y_1, y_2 , ?'

La relazione di meronimia/olonimia è transitiva (con le riserve che dopo verranno illustrate) e antisimmetrica e può anch'essa essere usata per costruire relazioni gerarchiche. Un esempio dell'applicazione del sistema gerarchico può essere il seguente: "becco" ed "ala" sono meronimi di "uccello", "canarino" è iponimo di "uccello". Quindi per il sistema gerarchico, "becco" e "ala" sono essi stessi anche meronimi di "canarino". L'applicazione della transitività invece ci porta a dire che se "dita" è meronimo di "mano" e "mano" è meronimo di "arto", allora "dita" è meronimo di "arto", ovvero le "dita" sono parte della "mano" e quindi anche dell'"arto". Questo non è sempre vero e dipende dal tipo di relazione "parte di" che si instaura fra le parti. Per chiarire meglio quanto affermato, ci si avvale di un esempio: "maniglia" è meronimo di "porta", "porta" è meronimo di "casa"; dire che "maniglia" è meronimo di "casa" non risulta certo molto appropriato. Risulterebbe singolare affermare che "maniglia" è parte di "casa". Da questo esempio, che illustra i paradossi nei quali si può incorrere pensando di applicare la transitività indistintamente ad ogni relazione di meronimia/olonimia, si intuisce che esistono varie tipologie di relazioni "parte di" o "parte/tutto"; per l'esattezza ne sono state individuate sei:

- **Componente/Oggetto:** i componenti sono parte integrante di un oggetto, possono essere separati da esso e rispetto allo stesso hanno una funzione

specifica ("gambe"/"sedia");

- **Membro/Insieme:** è caratterizzata dalla relazione di appartenenza presente in matematica e non solo; un membro è separabile dal tutto ("giocatore"/"squadra");
- **Porzione/Intero:** l'intero costituisce un tutto omogeneo di cui le porzioni fanno parte e possono esserne separate ("fetta"/"torta");
- **Materiale/Oggetto:** è la relazione fra il tutto e i materiali presenti in esso; i materiali non sono separabili e non hanno una funzione specifica rispetto all'oggetto ("fibra di carbonio"/"bicicletta");
- **Fase/Processo:** la fase indica un periodo temporale che fa parte dell'evoluzione del processo e non è da quest'ultimo separabile ("adolescenza"/"crescita");
- **Località/Area:** è una relazione spaziale fra due aree fisiche; ogni posto è dello stesso tipo dell'area e non può esserne separato ("Italia"/"Europa").

I tipi di meronimia presi in considerazione nello sviluppo dei nomi in MultiWordNet sono tre:

- **Componente/Oggetto**
- **Membro/Insieme**
- **Materiale/Oggetto**

Il primo di questi tipi è statisticamente il più frequente. Nei domini di interesse del MultiWordNet sulla *tutela delle acque*, sono state individuate le seguenti relazioni: *acqua* è meronimo di *ambiente* e *tutela delle acque* è iponimo di *acqua*.

3.3.5 Relazioni morfologiche

Una importante classe di relazioni lessicali sono quelle morfologiche fra forme di parola. Un esempio è quello del plurale dei nomi; se un utente inserisce la parola "trees", e lancia la ricerca, il programma non dovrebbe dare come risultato l'assenza della parola del database. Se la parola "trees" è in relazione morfologica con il suo singolare, la ricerca viene fatta sul termine "tree". Per evitare inutili ridondanze, nella costruzione di WordNet, gli implementatori si resero conto della necessità di introdurre nel software una parte che gestisse le relazioni morfologiche. Il lavoro non si presentò comunque così semplice, per le numerose forme irregolari di nomi e verbi.

4 Esempi di utilizzo del MultiWordNet specialistico per il dominio della tutela delle acque

La prima esigenza ravvisata per l'inserimento dei termini in MultiWordNet è stata quella di aggiungere ai domini presenti al momento, quelli di: *ambiente*, *acqua*, *tutela delle acque*. Contestualmente all'inserimento dei *synset*, sono state implementate le relazioni semantiche e lessicali precedentemente individuate; sono state inoltre inserite, per ogni *synset*, gli estremi delle normative di riferimento correlate, e, quando presenti, le definizioni. Nei paragrafi successivi vengono mostrate nel dettaglio le operazioni di ricerca dei termini e visualizzazione delle relazioni nel MultiWordNet sulla *tutela delle acque*.

4.1 Ricerca dei termini

Il database di MultiWordNet è consultabile all'indirizzo Internet:

<http://tcc.fbk.eu/projects/multiwordnet/>

previa autenticazione dell'utente tramite login e password personali.

4 ESEMPI DI UTILIZZO DEL MULTIWORDNET SPECIALISTICO PER IL DOMINIO DELLA TUTELA DELLE ACQUE

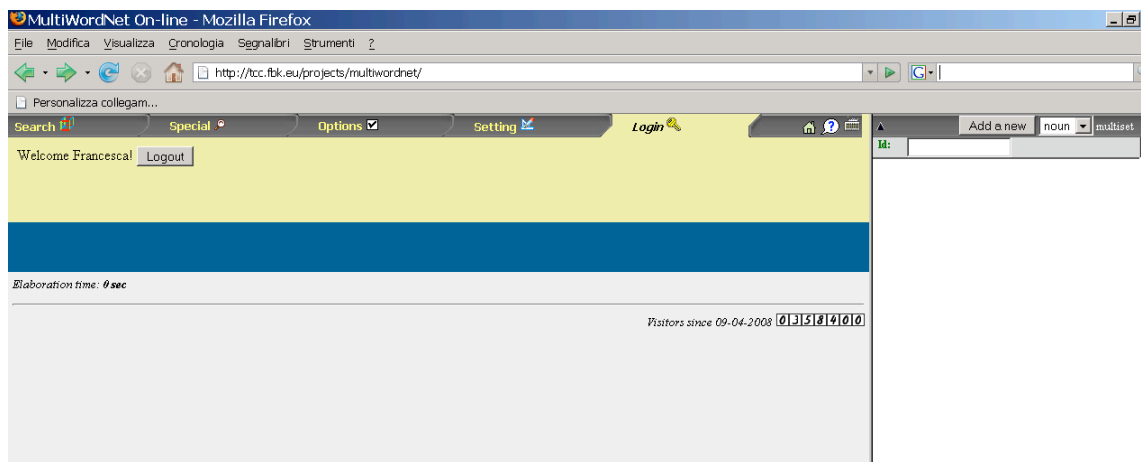


Figura 2: Schermata presentata da MultiWordNet ad autenticazione effettuata

Dopo che all'utente è stato consentito l'accesso, tramite login e password personali, come mostrato in Figura 2, il primo passo da compiere è il settaggio del browsing (consultazione del database senza apporre modifiche) e dell'editing (scrittura e modifica delle informazioni contenute nel database). Eseguendo il settaggio come in Figura 3, si potranno eseguire sia ricerche che modifiche nel MultiWordNet specialistico per la tutela delle acque.

4 ESEMPI DI UTILIZZO DEL MULTIWORDNET SPECIALISTICO PER IL DOMINIO DELLA TUTELA DELLE ACQUE

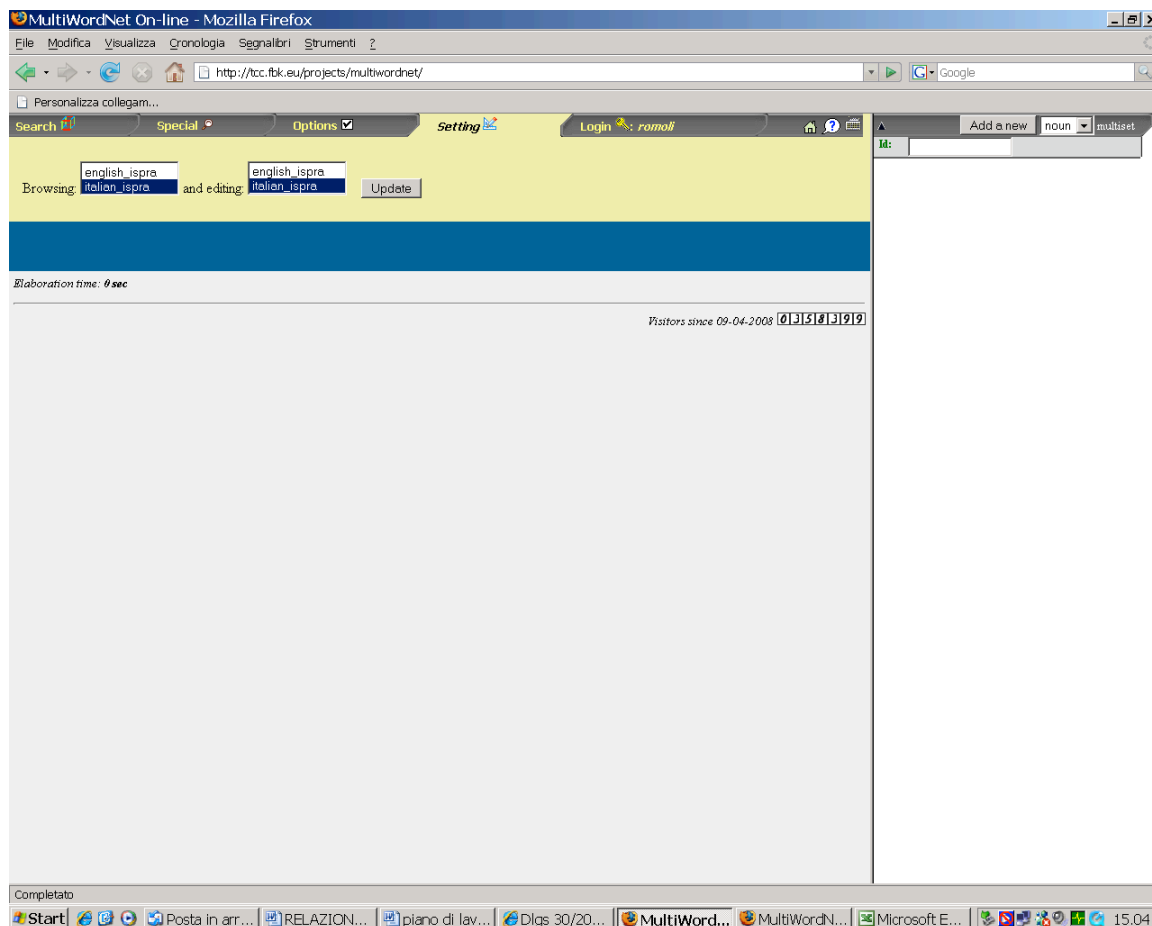


Figura 3: Settaggio di browsing ed editing in lingua italiana del MultiWordNet per la tutela delle acque

L'utente che voglia consultare il data base di MultiWordNet e lanciare ricerche di termini, deve selezionare l'etichetta "Search" dal pannello dei comandi che compare in alto nella parte sinistra della pagina Web. Comparirà quindi il menu di ricerca, come si vede in Figura 4. Il menù a tendina offre la possibilità di eseguire diversi tipi di ricerche, tra le quali:

- Ricerca di singole parole o termini, anche composti (Word)

- Ricerca di termini che contengono nella loro definizione determinate parole (Gloss)
- Ricerca di tutti i termini definiti in un determinato dominio (Domain)

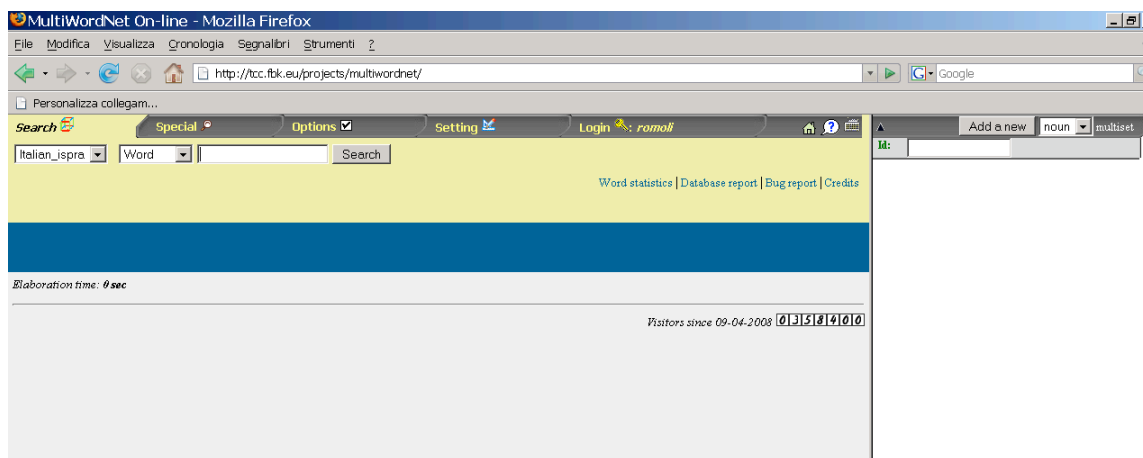


Figura 4: Schermata presentata da MultiWordNet a seguito della selezione da parte dell'utente dell'etichetta "Search"

In Figura 5 viene mostrata la schermata che presenta all'utente la possibilità di scelta del criterio di ricerca.

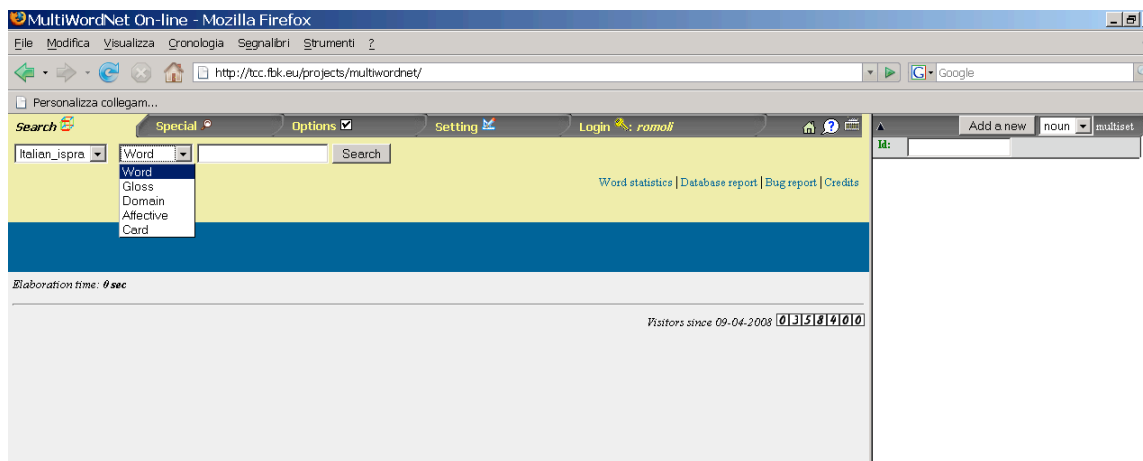


Figura 5: Scelta dei diversi criteri di ricerca in MultiWordNet

Di seguito, in Figura 6 viene presentato il risultato della ricerca della parola "acqua".

The screenshot shows the MultiWordNet On-line interface in a Mozilla Firefox browser. The search bar at the top contains the word "acqua". The results section, titled "The word 'acqua' has 8 senses:", lists eight different meanings of the word, each with a unique identifier and a brief description. The senses are categorized by domain: Ecology, Geography, Chemistry, Philosophy, Meteorology, Food, Jewellery, and Factotum. On the right side, a detailed view of the first sense (n#00005656) is shown, including its sense, position, domain, and a detailed gloss.

Sense	Domain	Description
1. (n#06666185) acqua	(Ecology, Earth, Geography)	massa, distesa d'acqua
2. (n#Q0005602) acqua, acqua	(Water_protection, Water, Environment)	
3. (n#10650211) acqua, H2O	(Chemistry, Geography)	liquido trasparente incolore, inodore e insapore, la cui molecola è formata da due atomi di idrogeno e uno di ossigeno, costituente fondamentale degli organismi viventi
4. (n#10651368) acqua	(Philosophy)	nella filosofia antica, uno dei quattro elementi di cui era composto l'universo
5. (n#07830260) pioggia, acqua	(Meteorology)	precipitazione di gocce d'acqua dal vapore condensato nell'atmosfera, "due giorni di pioggia senza interruzione", "sono uscito senza l'ombrello e ho preso l'acqua"
6. (n#W0021110) acqua	(Alimentation, Chemistry, Pharmacy)	nome generico di soluzioni acquose o alcoliche di preparazione alimentare, farmaceutica, chimica, ecc.; "acqua tonica, acqua di rose, acqua ossigenata, acqua ragia"
7. (n#W0021111) acqua	(Jewellery)	trasparenza, limpidezza, purezza, qualità di una pietra preziosa; "un diamante dall'acqua purissima"
8. (n#W0021112) acqua	(Factotum)	liquido presente in alcune piante e frutti; "fare cuocere a fuoco lento e quando l'acqua del pomodoro si sarà ritirata, togliere dal fuoco e passare al passaverdure"

Elaboration time: 0 sec

Visitors since 09-04-2008: 0131514100

Figura 6: Risultato della ricerca del termine "acqua"

Come si evince dall'immagine della pagina Web di ricerca, la parola "acqua" presenta otto diversi sensi o significati, che essa assume in altrettanti contesti tematici (domini) tra quelli inseriti nel database.

Nell'esempio che segue si fa riferimento alla ricerca del termine "acque sotterranee".

Come si può vedere in Figura 7 compaiono i domini ai quali il termine si riferisce: *water protection*, *water*, *environment*, la sua definizione e la normativa dalla quale il termine è stato estrapolato.

4 ESEMPI DI UTILIZZO DEL MULTIWORDNET SPECIALISTICO PER IL DOMINIO DELLA TUTELA DELLE ACQUE

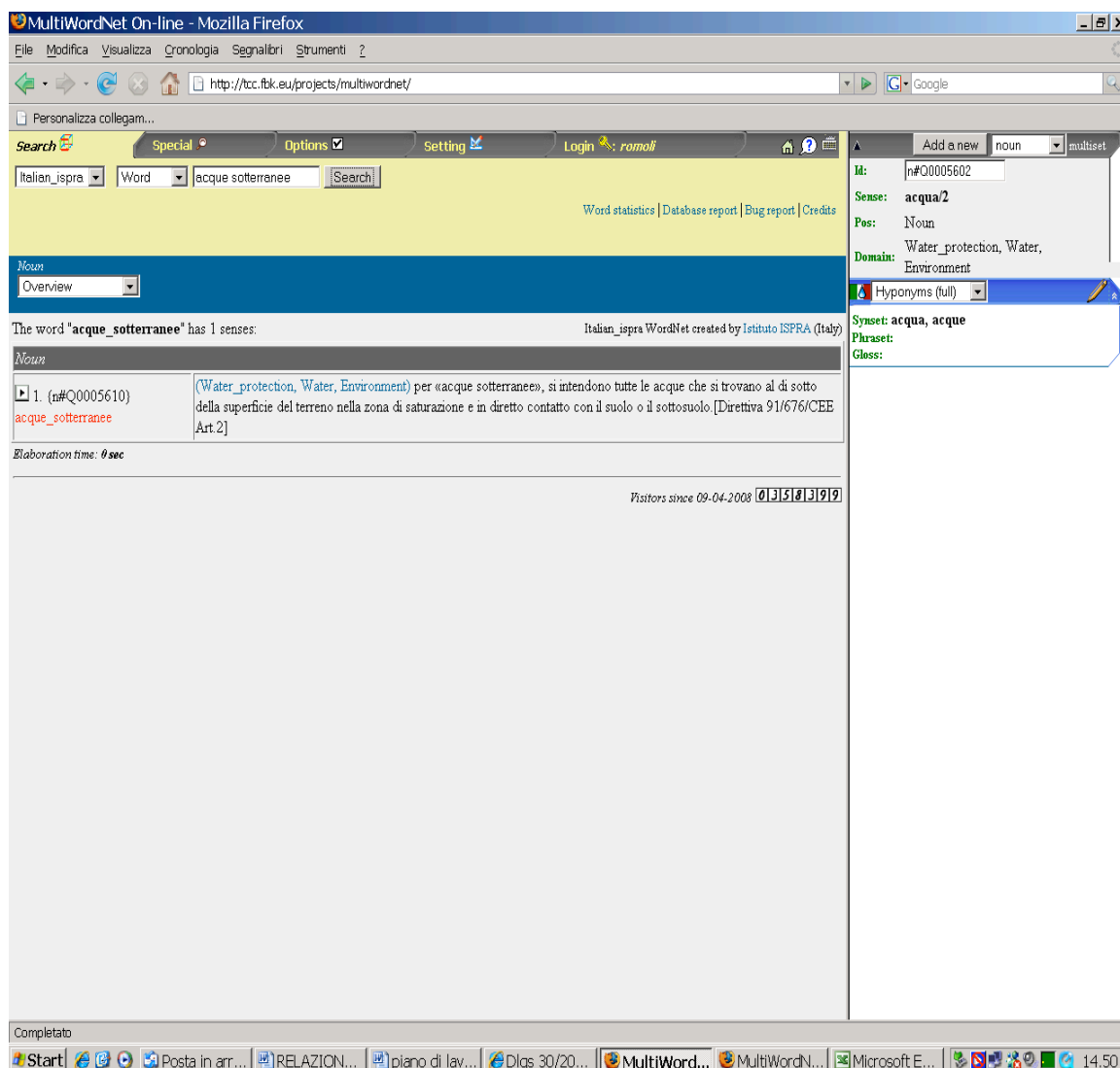


Figura 7: Risultato della ricerca del termine "acque sotterranee"

4.2 Visualizzazione delle relazioni tra termini

MultiWordNet, essendo basato su una "trama" di relazioni terminologiche, permette di navigare in modo completo nelle relazioni tra termini. Si faccia riferimento all'esempio del termine "acque sotterranee". Per conoscere i termini, presenti nel data

base, che gli sono relazionati, basta cliccare sulla finestra "Overview" presente sulla pagina web visualizzata, come mostrato in Figura 8. In questo caso si ha solo un iperonimo e cioè "acque" che in questo data base è stato reso sinonimo di "acqua".

4 ESEMPI DI UTILIZZO DEL MULTIWORDNET SPECIALISTICO PER IL DOMINIO DELLA TUTELA DELLE ACQUE

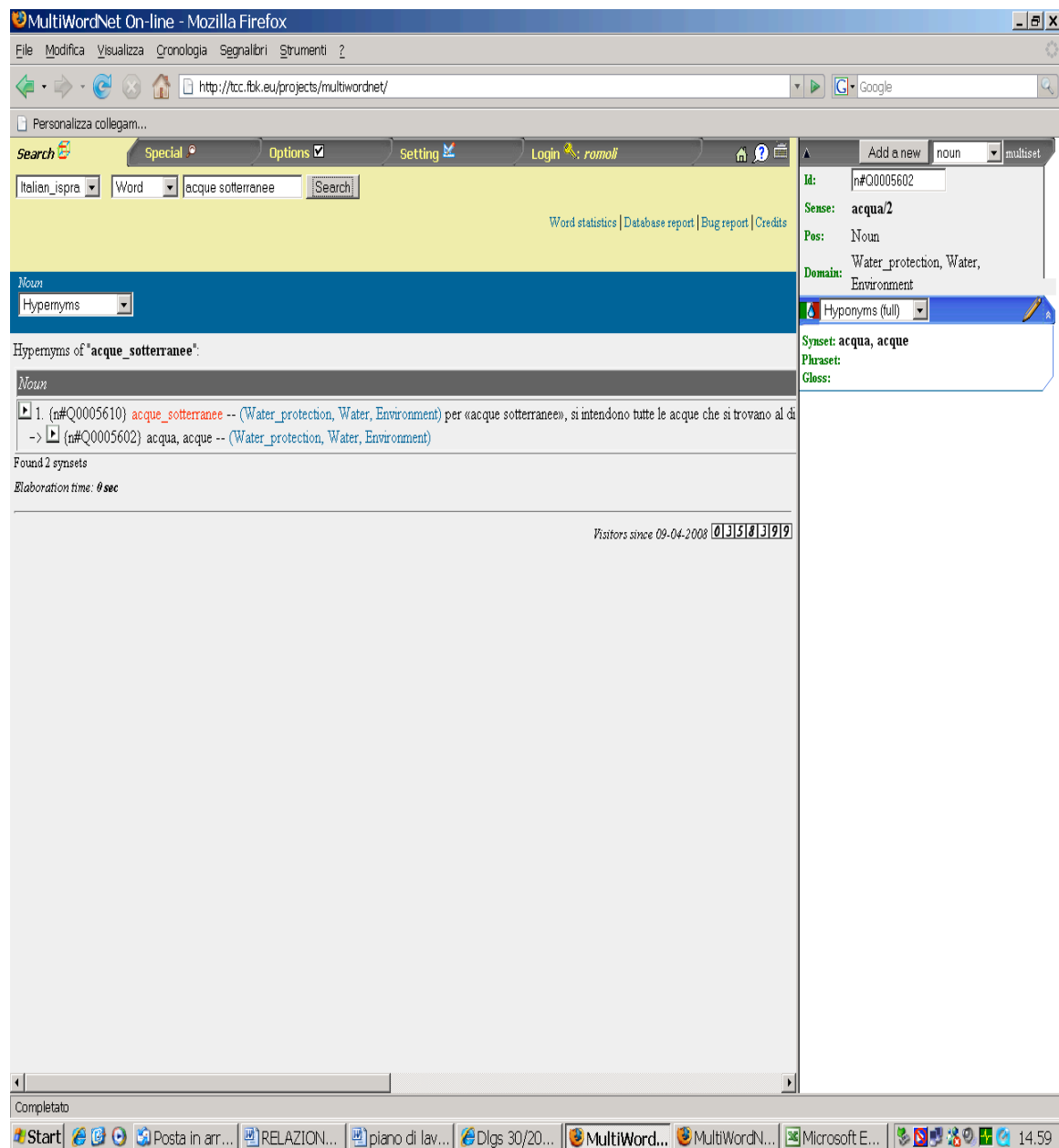


Figura 8: Visualizzazione degli iperonimi del termine "acque sotterranee"

In Figura 9 vengono mostrati i risultati della ricerca del termine "lago".

4 ESEMPI DI UTILIZZO DEL MULTIWORDNET SPECIALISTICO PER IL DOMINIO DELLA TUTELA DELLE ACQUE

The screenshot displays the MultiWordNet On-line interface in a Mozilla Firefox browser window. The address bar shows the URL <http://tcc.ftk.eu/projects/multiwordnet/>. The interface includes a search bar with the word "lago" entered, and a dropdown menu for the language set to "Italian_ispra". The search results for the word "lago" are displayed, showing two senses:

- Sense 1: (n#Q0005656) **lago, laghi**. Description: (**Water_protection, Water, Environment**) un corpo idrico naturale lentico, superficiale, interno, fermo, di acqua dolce, dotato di significativo bacino scolante. Non sono considerati ambienti lacustri tutti gli specchi d'acqua derivanti da attività estrattive, gli ambienti di transizione, quali sbarramenti fluviali tratti di corsi d'acqua in cui la corrente rallenta fino ad un tempo di ricambio inferiore ad una settimana e gli ambienti che mostrano processi di interrimento avanzati che si possono definire come zone umide; "D.M. 16-06-2008 n. 131 All. 1 Sezione A A.2.1 --Direttiva 91/271/CEE Allegato 2", "Direttiva 91/271/CEE Allegato 2".
- Sense 2: (n#06787575) **lago**. Description: (**Geography**) massa d'acqua circondata dalla terra.

The interface also shows a sidebar with various options and a bottom status bar indicating the completion of the search.

Figura 9: Risultato della ricerca del termine "lago"

Come si può vedere il termine "lago" nel data base è stato inserito in due tipologie diverse di domini e quindi ha due diverse definizioni. La prima riguarda i nostri domini di interesse e la normativa dalla quale è stata estrapolato, e rappresenta quindi la definizione del MultiWordNet della *tutela delle acque*. L'altra definizione del termine "lago" è generica e si riferisce al MultiWordNet in italiano della "Fondazione Bruno Kessler". Se richiesti si possono ottenere gli iponimi di "lago", presenti nel data base, con la relativa struttura gerarchica, come si vede in Figura 10.

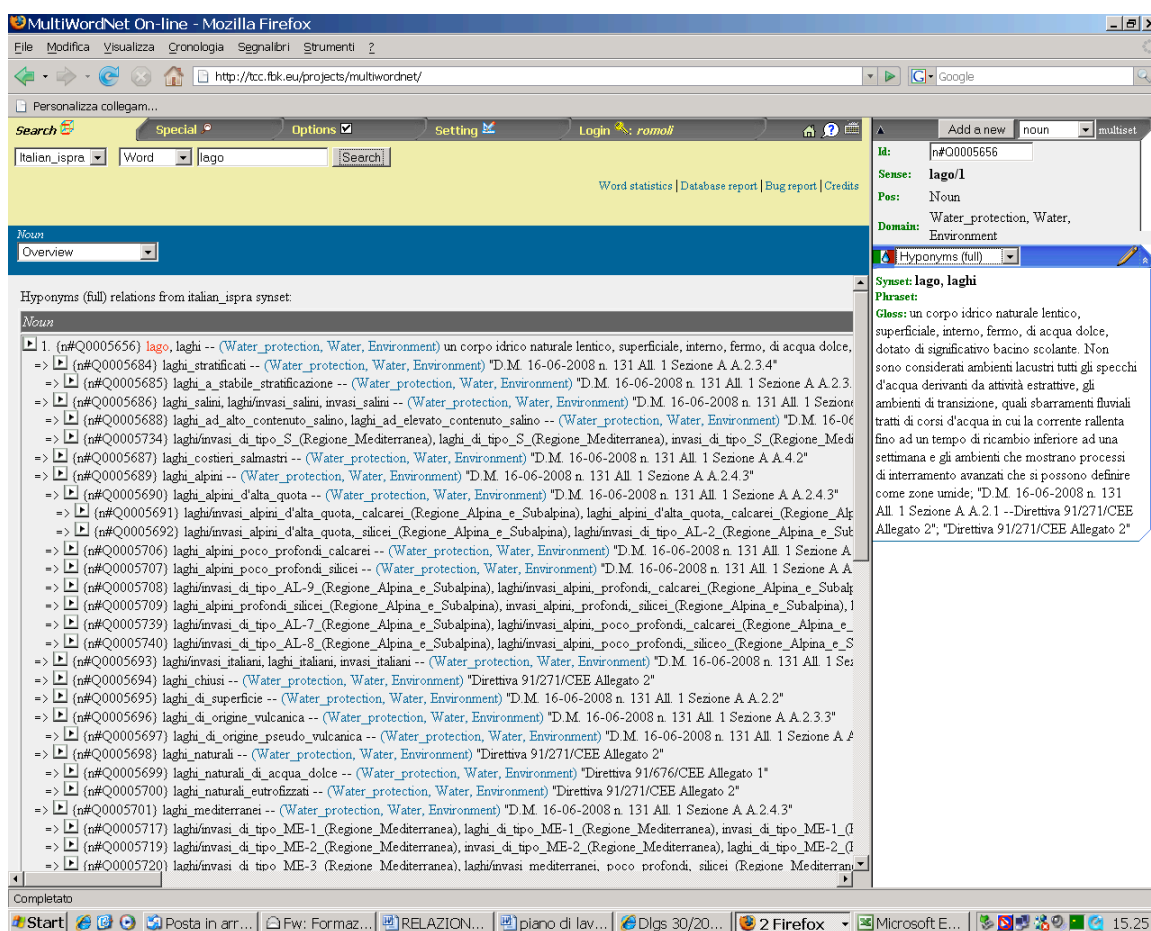


Figura 10: Risultato della ricerca degli iponimi del termine "lago"

Come si può vedere, il termine "lago" nel data base ha due sensi, in quanto, essendo stato inserito in due tipologie diverse di domini, presenta due diverse definizioni. La prima definizione riguarda i domini di interesse di questa tesi, cioè *acqua*, *ambiente* e *tutela delle acque* e contiene la normativa dalla quale è stata estrapolata. Per navigare all'interno di uno dei sensi della parola "lago" occorre cliccare sul quadratino grigio contenente la freccia nera, che si trova accanto alla parola scelta (nell'esempio in figura viene visualizzata anche la definizione della parola scelta, sul menu di destra della pagina Web), come mostrato in Figura 11.

The screenshot displays the MultiWordNet On-line interface. The search results for the word "lago" are as follows:

Sense	ID	Definition
1. (n#Q0005656) lago, laghi	n#Q0005656	(Water_protection, Water, Environment) un corpo idrico naturale lentic, superficiale, interno, fermo, di acqua dolce, dotato di significativo bacino scolante. Non sono considerati ambienti lacustri tutti gli specchi d'acqua derivanti da attività estrattive, gli ambienti di transizione, quali sbarramenti fluviali tratti di corsi d'acqua in cui la corrente rallenta fino ad un tempo di ricambio inferiore ad una settimana e gli ambienti che mostrano processi di interrimento avanzati che si possono definire come zone umide; "D.M. 16-06-2008 n. 131 All. 1 Sezione A A.2.1 --Direttiva 91/271/CEE Allegato 2", "Direttiva 91/271/CEE Allegato 2"
2. (n#06787575) lago	n#06787575	(Geography) massa d'acqua circondata dalla terra

The right sidebar provides additional details for the selected sense (lago, laghi):

- Sense:** lago/1
- Pos:** Noun
- Domain:** Water_protection, Water, Environment
- Hyponyms (full):** lago, laghi
- Phrasal:** un corpo idrico naturale lentic, superficiale, interno, fermo, di acqua dolce, dotato di significativo bacino scolante. Non sono considerati ambienti lacustri tutti gli specchi d'acqua derivanti da attività estrattive, gli ambienti di transizione, quali sbarramenti fluviali tratti di corsi d'acqua in cui la corrente rallenta fino ad un tempo di ricambio inferiore ad una settimana e gli ambienti che mostrano processi di interrimento avanzati che si possono definire come zone umide; "D.M. 16-06-2008 n. 131 All. 1 Sezione A A.2.1 --Direttiva 91/271/CEE Allegato 2", "Direttiva 91/271/CEE Allegato 2"

Figura 11: Risultato della selezione del primo senso del termine "lago"

Navigando all'interno del senso prescelto della parola, si possono inoltre ottenere i sinonimi, gli iponimi e gli iperonimi del termine "lago" presenti nel data base, con la relativa struttura gerarchica. Cliccando sul menu a tendina dell'icona contenente la matita, si ottiene la possibilità di visualizzare dei termini correlati col termine "lago", in funzione del tipo di relazione prescelta, come mostrato in Figura 12.

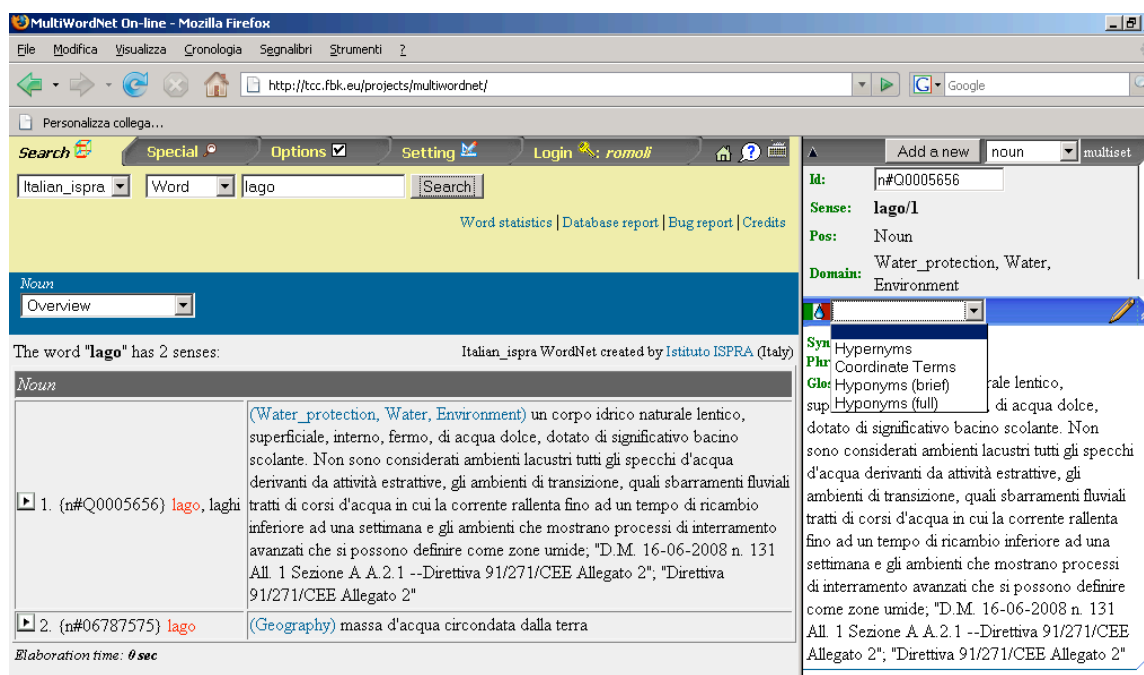


Figura 12: Menu a tendina per la visualizzazione dei termini in relazione col termine "lago"

Se si sceglie Hyponyms (full), si otterrà la lista di tutti gli iponimi del termine "lago", così come sono stati definiti nella fase di inserimento, espansi fino al massimo livello di dettaglio. Nel caso trattato nell'esempio, si otterrà la schermata di Figura 13.



In figura viene mostrata la lista degli iponimi del termine "lago" in scala gerarchica: ad esempio, "lago" ha, tra gli altri iponimi, "laghi stratificati", il quale a sua volta ha come iponimo il termine "laghi a stabile stratificazione".

Se si desidera visualizzare in dettaglio uno dei termini presenti nell'elenco degli iponimi di "lago", ad esempio "lago polimittico", occorre cliccare sul quadratino grigio contenente la freccia nera accanto al termine stesso. Sul menu a destra verrà visualizzata un'area contenente la definizione e la normativa dalla quale è stata estrapolata, come mostrato in Figura 14.

The screenshot shows the MultiWordNet On-line interface in a Mozilla Firefox browser. The search bar contains "lago" and the results are displayed in a list. The term "lago polimittico" is highlighted, and a detailed view is shown on the right. The detailed view includes the following information:

- Synset:** lago_polimittico
- Id:** n#Q0005702
- Pos:** Noun
- Domain:** Water_protection, Water, Environment
- Phrasal:**
- Gloss:** un lago/invaso è definito polimittico se non mostra una stratificazione termica evidente e stabile. Un ambiente lentic di questo genere pu_ andare incontro a diverse fasi di mescolamento nel corso del suo ciclo annuale. Per distinguere i laghi/invasi polimittici da quelli a stabile stratificazione vengono identificati i seguenti valori di profondità media:- < 3 m per i laghi/invasi al di sotto di 2000 m s.l.m.; - < 5 m per i laghi/invasi al di sopra di 2000 m s.l.m.; "D.M. 16-06-2008 n. 131 All. 1 Sezione A A.2.3"

Figura 14: Visualizzazione del dettaglio del termine "lago polimittico", iponimo di "lago"

Se inoltre, si desidera consultare la scheda completa relativa al termine prescelto, si deve cliccare sull'icona della matita gialla presente sulla destra della pagina e si otterrà la schermata mostrata in Figura 15.

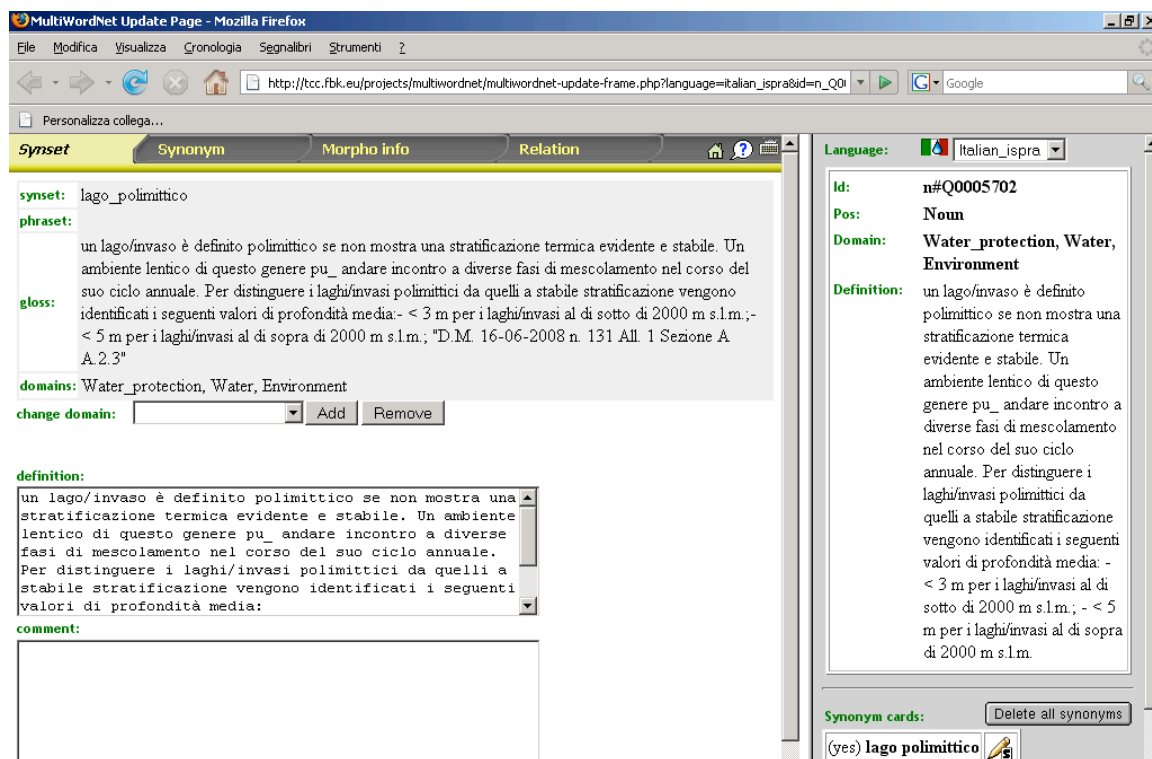


Figura 15: Visualizzazione della scheda relativa al termine "lago polimittico", iponimo di "lago"

È possibile anche interrogare MultiWordNet per ricercare gli iperonimi di un dato termine. Nell'esempio seguente, la parola di cui si vogliono ottenere gli iperonimi è "laghi naturali eutrofizzati". Per prima cosa si procede alla ricerca del termine prescelto, come mostrato in Figura 16.

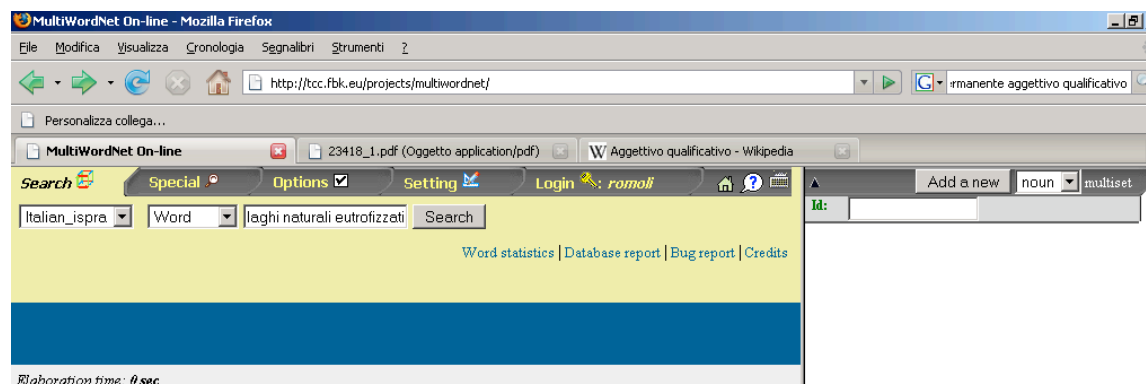


Figura 16: Ricerca del termine "laghi naturali eutrofizzati"

Cliccando sul bottone Search, MultiWordNet restituisce un solo significato o senso del termine "laghi naturali eutrofizzati", come mostrato in Figura 17.

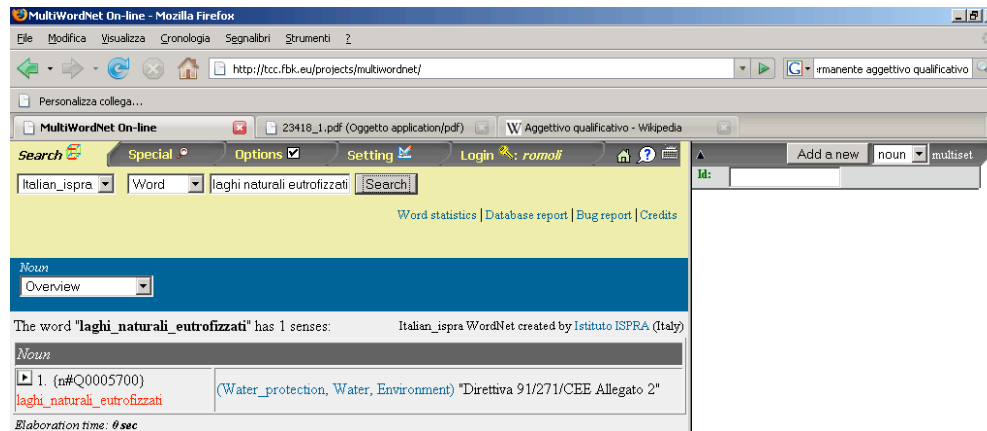


Figura 17: Il termine "laghi naturali eutrofizzati" presenta un solo senso

Per visualizzare gli iperonimi del termine "laghi naturali eutrofizzati", si deve cliccare sul bottone grigio contenente la freccia nera accanto al termine stesso; sul menu a destra verrà visualizzata un'area contenente la normativa da cui è stato estratto il termine, come mostrato in Figura 18.

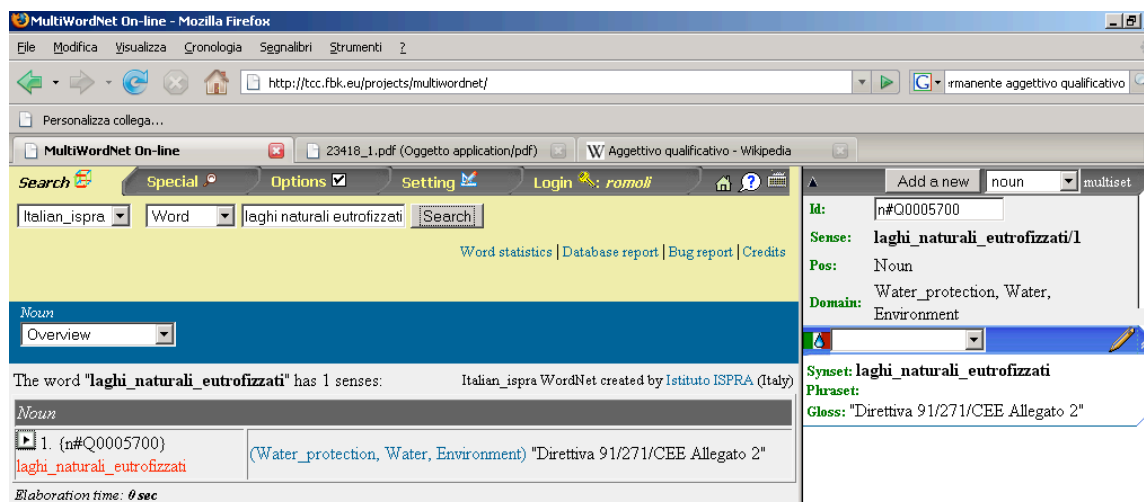


Figura 18: Risultato della selezione del primo e unico senso del termine "laghi naturali eutrofizzati"

Cliccando sul menu a tendina dell'icona contenente la matita, si ottiene la possibilità di visualizzare gli iperonimi del termine in questione.

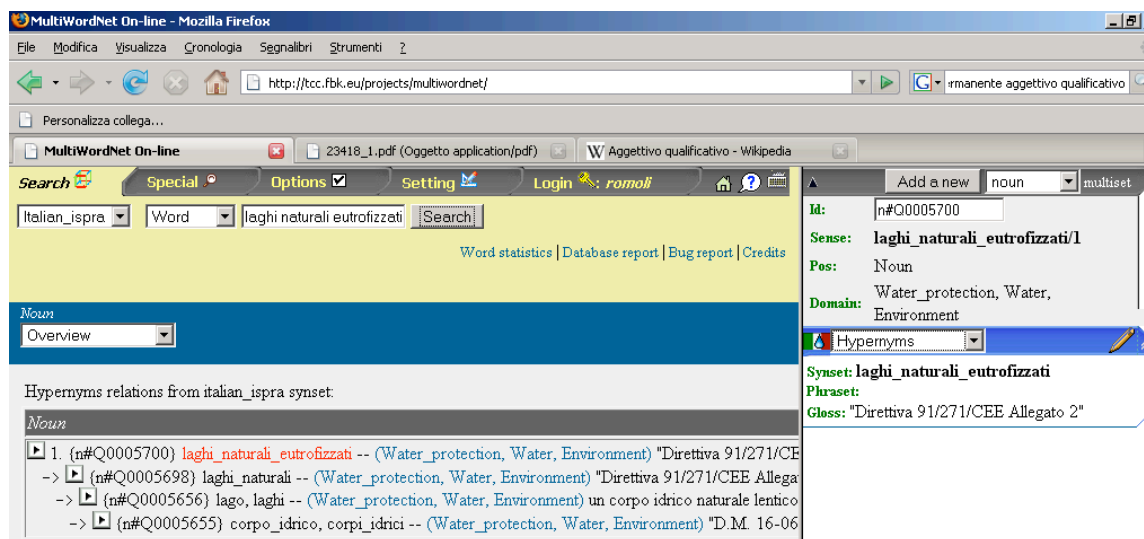


Figura 19: Visualizzazione della lista di iperonimi del termine "laghi naturali eutrofizzati"

In Figura 19 viene mostrata la lista degli iperonimi del termine "laghi naturali eutrofizzati" in scala gerarchica: "laghi naturali eutrofizzati" ha come iperonimo "laghi naturali", che a sua volta ha come iperonimo "lago/laghi", che a sua volta ha come iperonimo "corpo idrico/corpi idrici".

MultiWordNet è uno strumento che si basa sulle relazioni sintattiche e semantiche esistenti tra i termini; negli esempi seguenti verranno mostrati degli esempi di visualizzazione delle relazioni tra i termini, in particolare per quel che riguarda l'iponimia/iperonimia, la sinonimia e l'antonimia. In Figura 20 vengono mostrati gli iponimi del termine "acqua" presenti allo stato attuale nel dizionario specialistico:

4 ESEMPI DI UTILIZZO DEL MULTIWORDNET SPECIALISTICO PER IL DOMINIO DELLA TUTELA DELLE ACQUE

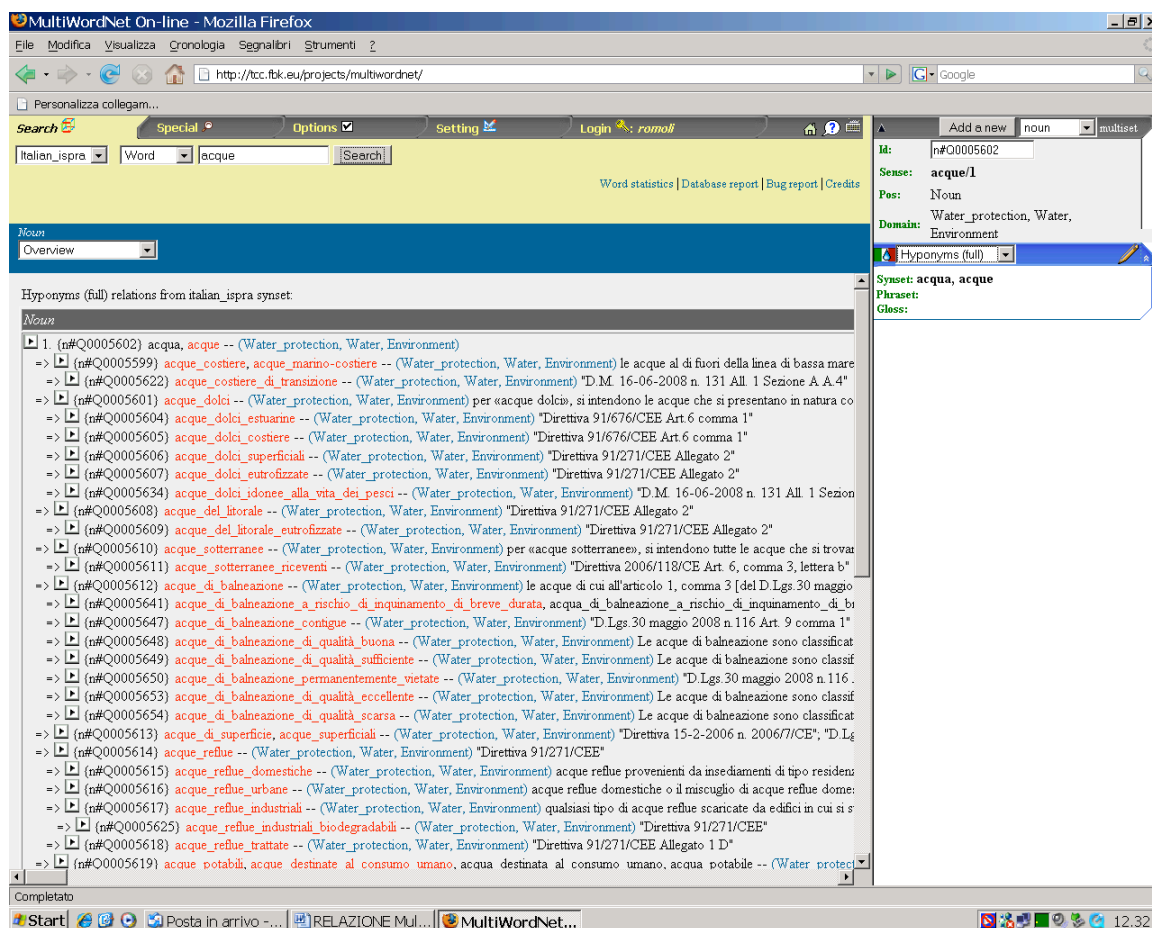


Figura 20: Visualizzazione in MultiWordNet degli iponimi inseriti del termine "acqua"

Come già accennato nella descrizione della struttura del database WordNet, i termini vengono in esso organizzati in synset, ossia insiemi di sinonimi. Nell’inserimento dei termini in MultiWordNet, sono state rilevate all’interno dei nostri domini di interesse, alcune relazioni di sinonimia; una di queste riguarda i due termini: ”acque costiere” e ”acque marino-costiere”. Per visualizzarla basta effettuare la ricerca di uno dei due termini, ad esempio ”acque costiere”.

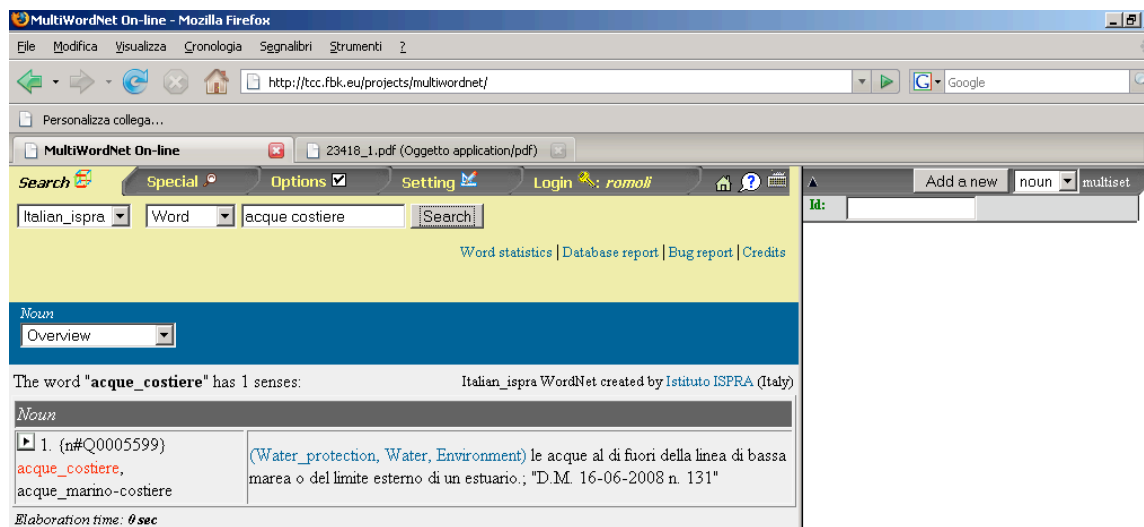


Figura 21: Visualizzazione in MultiWordNet dei sinonimi inseriti del termine ”acque costiere”

Come si vede in Figura 21, l'esito della ricerca restituisce immediatamente il termine cercato e il suo sinonimo "acque marino-costiere". In Figura 22 viene presentato un ulteriore esempio di sinonimia; vengono mostrati i sinonimi del termine "acqua potabile".

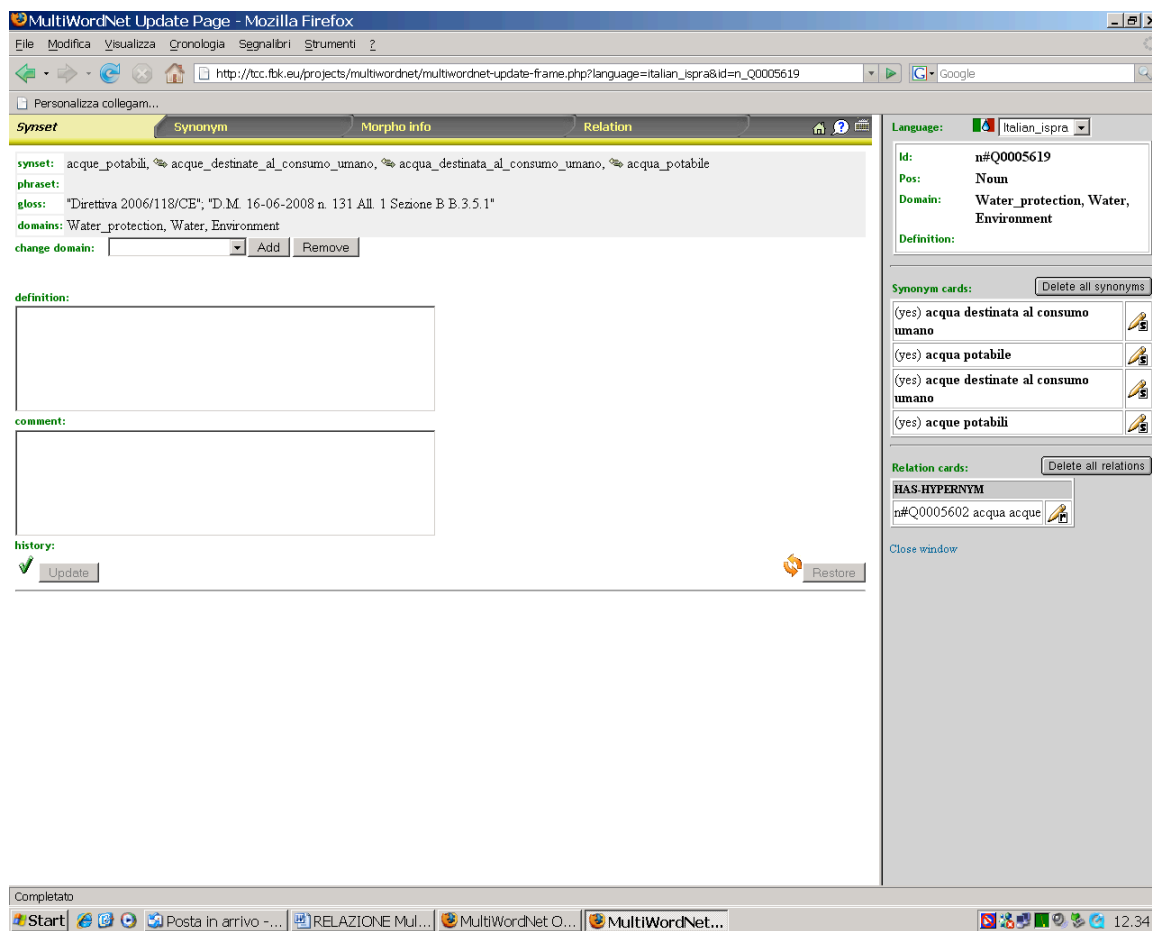


Figura 22: Visualizzazione in MultiWordNet dei sinonimi inseriti del termine "acqua potabile"

Nell’inserimento dei termini in MultiWordNet, sono state rilevate all’interno dei nostri domini di interesse, alcune relazioni di antonimia; una di queste riguarda i due termini: "acque superficiali" e "acque sotterranee". Per visualizzarla basta effettuare la ricerca di uno dei due termini, ad esempio "acque superficiali". In Figura 23 viene mostrato il risultato di tale ricerca.

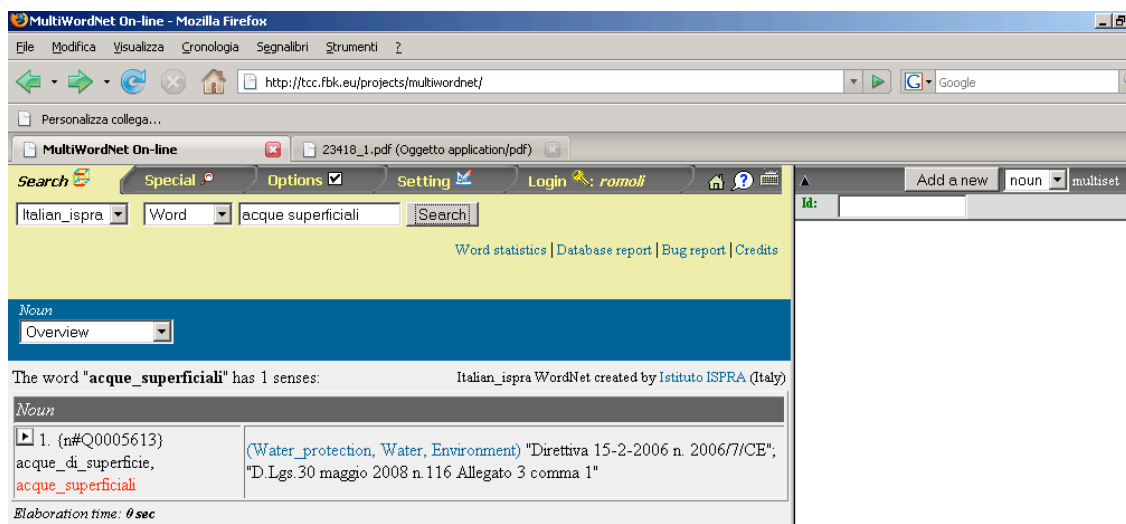


Figura 23: Ricerca del termine "acque superficiali"

Come si evince dalla Figura 23, il termine "acque superficiali" presenta anche una relazione di sinonimia col termine "acque di superficie". Per visualizzare il suo antonimo, basta cliccare sul bottone grigio contenente la freccia nera accanto al termine stesso, come mostrato in Figura 24.

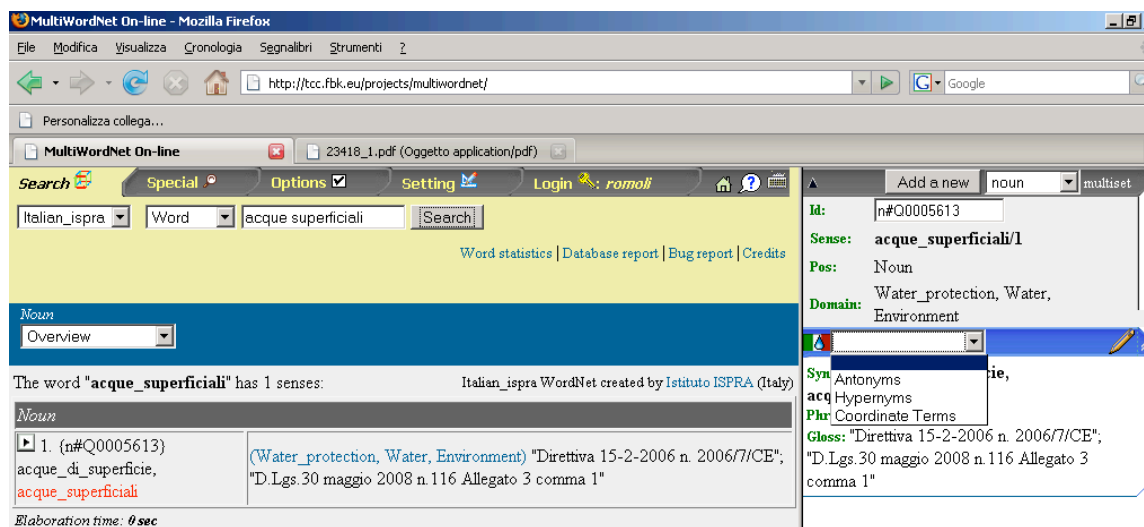


Figura 24: Visualizzazione delle relazioni del termine "acque superficiali"

Cliccando sul menu a tendina dell'icona contenente la matita, si ottiene la possibilità di visualizzare gli antonimi del termine "acque superficiali".

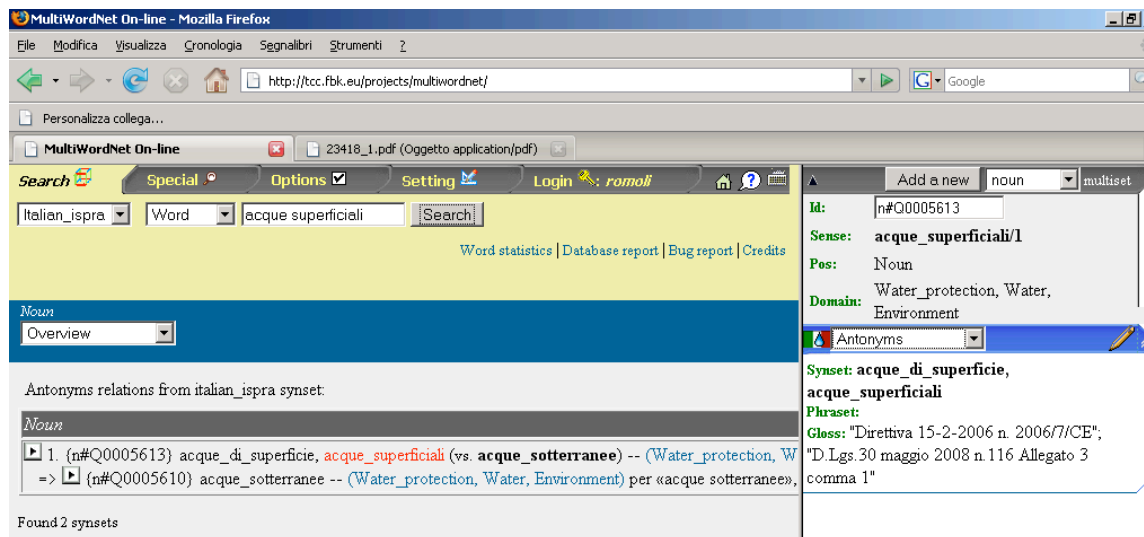


Figura 25: Visualizzazione degli antonimi inseriti del termine "acque superficiali"

Come si può vedere in Figura 25, il synset contenente i termini "acque superficiali" e "acque di superficie" ha come antonimo, nel dominio di interesse, il termine "acque sotterranee". Come ulteriore esempio di ricerca di antonimo, e a riprova della transitività della relazione di antonimia, in Figura 26 si presenta il risultato della ricerca dell'antonimo inserito nel vocabolario specializzato del termine "acque sotterranee", cioè il termine "acque superficiali".

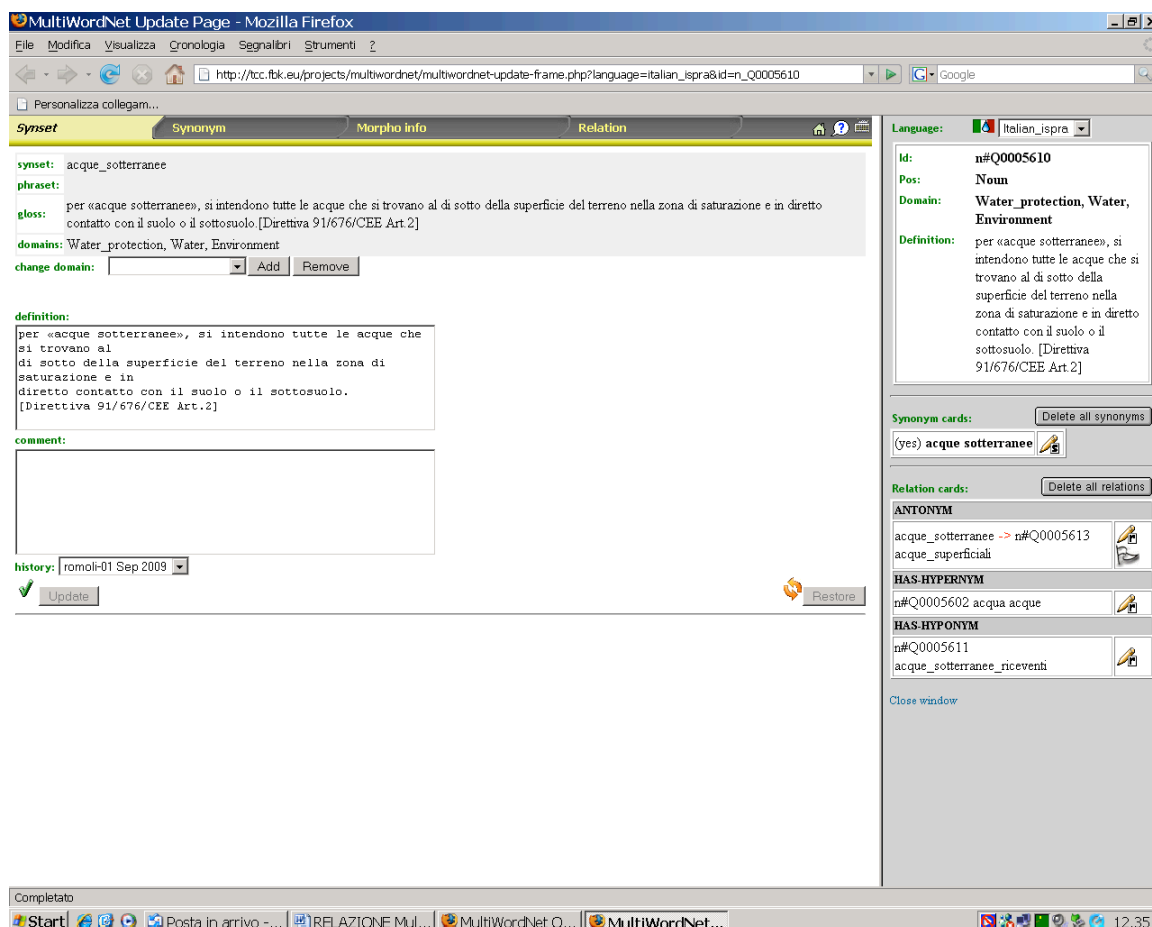


Figura 26: Visualizzazione in MultiWordNet dell'antonimo inserito del termine "acque sotterranee"

5 Conclusioni e sviluppi futuri

Lo strumento MultiWordNet, benchè ancora in fase di perfezionamento, si è rivelato molto potente e flessibile; ha permesso infatti di indagare a fondo la struttura della terminologia sulla *tutela delle acque* e di creare un dizionario specialistico completo sulla *tutela delle acque*. Il dizionario risultante è di pronta ed agile consultazione da parte dell'utenza a tutti i livelli; esso infatti, contiene sia le informazioni immediate sul termine (definizione ove presente, normative nelle quali il termine è presente) che quelle di livello più alto a cui sono interessati utenti più esperti (relazioni tra termini). Il lavoro di inserimento dei termini ha visto fino ad oggi una popolazione di circa 400 termini con relative relazioni, definizioni ove presenti e riferimenti normativi; nel futuro si prevede di continuare il lavoro di inserimento termini per la lingua italiana e di iniziare lo stesso lavoro di analisi terminologica ed inserimento termini, anche per la lingua inglese e la lingua spagnola. Occorrerà anche mantenere il dizionario e provvedere agli aggiornamenti che le normative recenti impongono, al fine di ottenere un prodotto affidabile, aggiornato e pienamente fruibile dall'utenza che sempre più a diversi livelli, è oggi interessata alla *tutela delle acque*.

6 Ringraziamenti

Ringrazio:

il mio tutor, il prof. Attilio Colagrossi

la mia co-tutor e compagna d'avventura, la dott.ssa Annamaria Rizzo

i ricercatori della Fondazione Bruno Kessler di Trento: Emanuele Pianta, Manuela Speranza, Christian Girardi

Riferimenti bibliografici

- [1] Emanuele Pianta, Luisa Bentivogli, and Christian Girardi, *MultiWordNet: Developing and Aligned Multilingual Database*.
- [2] C. Felbaum, *WordNet: An electronic lexical database*, 1998.
- [3] G. A. Miller, *WordNet: An on-line lexical database*, 1990.
- [4] Bernardo Magnini and Carlo Strapparava, *Costruzione di una base di conoscenza lessicale per l'italiano basata su WordNet*.