

UNIVERSITÀ DEGLI STUDI DI PADOVA  
FACOLTÀ DI SCIENZE STATISTICHE

CORSO DI LAUREA IN  
STATISTICA ECONOMIA E FINANZA



“STAGE PRESSO AXIA FINANCIAL RESEARCH:  
UTILIZZO DELLA PIATTAFORMA SYNTHEMA 2.0 PER L’ANALISI SEMANTICO-STATISTICA DEL WEB”

RELATORE: FRANCESCO LISI

LAUREANDO: GIOVANNI ALESSANDRO LAZZAROTTO

MATRICOLA: 544343-SEF

ANNO ACCADEMICO 2010-2011



## Sommario

<b>1. INTRODUZIONE .....</b>	<b>4</b>
<b>2. IL SOGGETTO OSPITANTE .....</b>	<b>5</b>
2.1 Gruppo Axia .....	5
2.2 Unità organizzativa.....	6
<b>3. ATTIVITA SVOLTE .....</b>	<b>7</b>
3.1 Borse Elettriche .....	7
3.1.1 Esempio: Powernext .....	8
3.2 Agenda Geo economica.....	9
3.2.1 Esempio .....	9
3.3 Petrolio e commodities .....	11
3.3.1 Esempio: commento .....	12
<b>4. APPROFONDIMENTO: SYNTHEMA 2.0 .....</b>	<b>14</b>
4.1 Cos'è Synthema 2.0.....	14
4.2 Come funziona Synthema 2.0.....	14
4.2.1 Fase 1: raccolta dati .....	14
4.2.2 Fase 2: base di conoscenza .....	15
4.2.3 Fase 3: analisi dei dati.....	15
4.3 Esempio: commenti degli utenti della rete sul risultato sportivo della Champions League 2009/2010.	15
4.3.1 Raccolta dati .....	15
4.3.2 Perfezionamento base di conoscenza.....	16
4.3.3 Elaborazione ed analisi dei dati .....	17
4.3.3.1 Grafici relazionali .....	17
4.3.3.2 Distribuzione temporale dei principali termini di interesse.....	21
4.3.3.3 Grafici a torta per i termini di interesse .....	22
4.3.4 Conclusioni .....	24
<b>5. SITOGRAFIA.....</b>	<b>26</b>

# 1 INTRODUZIONE

Nel corso degli ultimi anni, grazie soprattutto ai blog e ai social network, è diventata abitudine comune commentare e discutere su internet i principali argomenti di cronaca, politica, sport, musica, ecc. Ai fini di non trascurare questo enorme numero di dati è nato il bisogno di trovare un modo per trarne delle informazioni quantitative e qualitative.

La piattaforma Synthema 2.0 è nata con questi scopi, e nei miei tre mesi di stage presso Axia Financial Research ho avuto modo di utilizzarla per soddisfare la richiesta di questo tipo di conoscenze da parte di alcuni clienti.

Nel mio rapporto di lavoro con Axia ho anche svolto alcune mansioni di supporto alle pubblicazioni giornaliere e settimanali, quali il monitoraggio del mercato dell'energia delle principali Borse Elettriche Europee, del mercato delle commodities e dei dati macroeconomici.

Nella prima parte della relazione di stage ho voluto presentare il gruppo Axia e le mansioni che quotidianamente andavo a svolgere.

Nel capitolo 4 si trova invece un approfondimento sull'utilizzo della piattaforma Synthema 2.0 ai fini di comprendere quale sia stato l'impatto sul web della vittoria dell'F.C. Internazionale della Champions League 2009/2010.

## 2 IL SOGGETTO OSPITANTE

### 2.1 GRUPPO AXIA

Axia viene costituita nel 1994 a Padova, in forma di Società in accomandita semplice, dall'economista Fabrizio Spagna. L'obiettivo di Axia è promuovere l'integrazione nelle leggi del mercato dell'aspetto etico, promuovendo una cultura dell'economia dell'accesso alla globalizzazione e orientare al meglio lo sviluppo dell'azienda e gli investimenti. Il progetto strategico si sviluppa attraverso una struttura societaria che tende a coinvolgere, a vari livelli, più soggetti economici.

Axia Financial Research, con sede a Padova in Galleria Porte Contarine, 4, opera come advisor indipendente in ambito economico-finanziario ed etico. La società collabora con la società di investimento mobiliare Classica S.p.a. nella stesura di report giornalieri relativi ai principali indici borsistici italiani ed internazionali ed al mercato Forex oltre che alla produzione di report settimanali sul mercato delle commodities con particolare focus su petrolio, gas e mercato dell'energia. Il ramo "financial research" si occupa della redazione di report su società quotate alla borsa di Milano, con la finalità di cogliere le migliori opportunità d'investimento, i quali vengono pubblicati sul sito [www.borsaitaliana.it](http://www.borsaitaliana.it) e su alcune piattaforme ( Capital IQ, Reuters, Factset, Bloomberg), oltre che sul portale [www.axia.org](http://www.axia.org) alla sezione "Report Research". Axia Financial Research inoltre produce con scadenza quindicinale un articolo contenente la valutazione di una società operante nel settore energetico che viene pubblicata su Quotidiano Energia, una testata specializzata del settore energetico. Il report Axia destinato a Quotidiano Energia viene pubblicato sia nella versione cartacea del giornale che nel sito [www.quotidianoenergia.it](http://www.quotidianoenergia.it) all'interno della sezione "Energia a Piazza Affari", accessibile tramite abbonamento.

Axia si occupa anche di finanza etica, offre consulenza per la progettazione di portafogli etici, il monitoraggio e rating di fondi etici e di società, segue con costanza l'evolversi di questi temi nel mondo.

Nei settori delle utilities e telecomunicazioni, biotecnologie ed e-media il gruppo offre consulenza nella definizione di business plan e di progetti di finanziamento, effettua valutazioni della struttura finanziaria, analisi del mercato potenziale e analisi del rischio.

Occasionalmente vengono compiono studi e ricerche per investitori istituzionali e privati, a dealing rooms d'Istituti di credito, intermediari finanziari e fondi d'investimento, precise indicazioni sulla tendenza dei principali mercati finanziari, ottenuta principalmente attraverso l'analisi tecnica e sulla volatilità. Si forniscono scenari prospettici su valute, commodity, equity, interest rate su orizzonti temporali medio lunghi (3-6-12 mesi) attraverso modelli econometrici.

La struttura organizzativa del gruppo è semplice, con pochi livelli gerarchici, date anche le ridotte dimensioni. Nella sede operativa di Padova lavorano principalmente gli analisti finanziari, mentre nella sede di Roma si svolgono le attività inerenti la redazione, le relazioni esterne e l'amministrazione del gruppo. Una caratteristica di Axia risiede nel tentativo d'integrare gli studi economico-finanziari, l'aspetto della responsabilità sociale usufruendo dell'insieme informativo creato dalla ricerca giornalistica.

Il processo decisionale si concentra ai livelli individuabili dalle figure del "Capo analista" o "Capo redattore" che riferiscono al Direttore Responsabile cui spetta l'ultima decisione. Rimane il rischio con questo modello organizzativo di sovraccarico al vertice in cui si possono accumulare problemi irrisolti e dal quale può derivare una perdita del controllo delle attività in crescita.

## **2.2 UNITA' ORGANIZZATIVA**

Gli analisti finanziari della sede di Padova si occupano del report settimanale sul mercato del petrolio, gas, titoli energetici, oltre a redarre una nota giornaliera, in collaborazione con la Sim Classica, contenente il quadro operativo delle borse internazionali ed analisi macroeconomiche. A queste attività si affianca la ricerca di opportunità di investimento nel mercato borsistico italiano. Qualora si ritenga sottovalutata una società alla quotazione attuale di mercato si provvede alla redazione di un report avente come finalità la determinazione di un target price per la società in questione. Tali report vengono inizialmente inviati agli investor relator della società in esame per avere un feedback sul lavoro effettuato ed eventualmente apportare delle correzioni alla valutazione qualora, dopo il dialogo con la società in esame, si ritenga opportuno apportare modifiche alle ipotesi alla base del modello.

Gli analisti finanziari presenti nell'ufficio hanno mansioni con obiettivi ben specificati, ma con compiti flessibili perché spesso interagiscono tra loro nello svolgerli. L'ambiente di lavoro, due scrivanie una di fronte all'altra in un'unica grande stanza, favorisce la comunicazione informale, diretta e veloce. Capita che nascano gruppi di lavoro spontanei come risposta a particolari richieste del capo analista o dei clienti.

### **3. ATTIVITA' SVOLTE**

Ho iniziato la mia collaborazione come stagista in Axia in data Lunedì 12 Aprile 2010, nella sede operativa di Axia Financial Research situata in Galleria Porte Contarine n°4 a Padova.

Nel primo periodo dei due mesi trascorso in Axia ho svolto un ruolo di supporto alle pubblicazioni giornaliere e settimanali che vengono fatte, occupandomi in particolare delle Borse Elettriche Europee, del commento di dati macroeconomici, e del mercato delle commodity.

Nella seconda parte dello stage, inoltre, mi è stato affidato il ruolo di ricercatore semantico e statistico per l'analisi econometrica statistica del web: si tratta di utilizzare la piattaforma Synthema 2.0 per la valutazione della reputazione online e del posizionamento strategico di un'impresa.

Di seguito andrò a descrivere le mie mansioni di supporto alle pubblicazioni, mentre nella seconda parte della relazione farò un esempio dello svolgimento del lavoro con Synthema.

#### **3.1 BORSE ELETTRICHE**

Nell'ambito delle borse elettriche mi è stato affidato il ruolo di monitoraggio del mercato elettrico, con particolare attenzione alle principali borse elettriche europee. Ogni settimana avevo il compito di reperire i dati sul prezzo medio dell'energia e i volumi scambiati al mercato elettrico italiano ed estero ed inserirli in appositi file Excell. Per il venerdì, inoltre, dovevo procedere alla stesura di un articolo, da sottoporre ai membri dello staff di Axia che si occupavano delle pubblicazioni, sull'andamento delle borse elettriche.

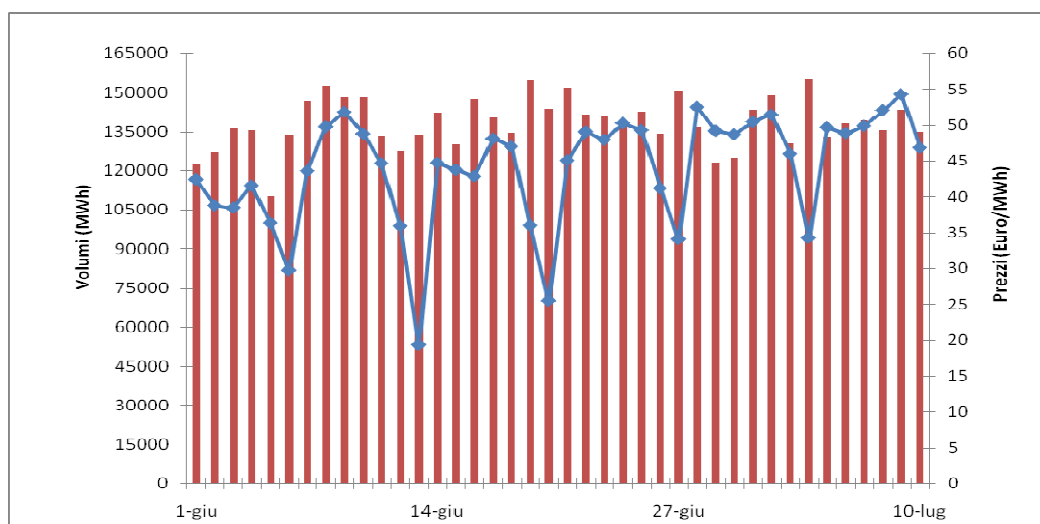
Per l'acquisto e la vendita di energia elettrica si utilizza un sistema organizzato di domanda e offerta chiamato appunto borsa elettrica. Al contrario delle normali borse valori, il prezzo di vendita dell'energia viene deciso il giorno prima, su base oraria, attraverso il sistema del prezzo marginale, o system marginal price. Tale procedimento, adottato nelle principali borse elettriche, consente di fissare il prezzo dell'energia elettrica ad un livello che è pari al prezzo dell'offerta più onerosa tra quelle accettate per soddisfare la domanda: è un meccanismo utile a corrispondere il prezzo di equilibrio tra domanda e offerta ai produttori.

Per monitorare l'andamento del prezzo medio dell'energia e il volume degli scambi si possono liberamente reperire i dati dal web. I mercati in esame sono quello italiano (Ipx), olandese (APX Power NL), francese (Powernext), spagnolo (Omel), inglese (APX Power UK), tedesco (EEX) e scandinavo (Nord Pool). La realizzazione dell'articolo non è altro che il confronto tra le medie settimanali (da domenica a sabato) del volume e dei prezzi medi giornalieri rispetto alle rilevazioni

precedenti: le quotazioni, al contrario della consueta borsa valori, sono riferite al giorno successivo e avvengono per tutti i giorni della settimana, 365 giorni l'anno. Ad inizio mese, inoltre, si procede con un confronto su base mensile e se opportuno su base trimestrale e quadrimestrale, ma ancora più importante è il confronto congiunturale, ovvero con i dati dell'anno prima.

### **3.1.1 ESEMPIO: POWERNEXT**

Di seguito è riportato il grafico, la tabella e i commenti per la settimana da domenica 4 a sabato 10 luglio del mercato elettrico francese del Powernext.



<b>Data</b>	<i>Max</i> (€/MWh)	<i>Min</i> (€/MWh)	<i>Average</i> (€/MWh)	<i>Volumi</i> (MWh)
4-lug	48.32	15.19	34.286	155172
5-lug	68.25	27.88	49.725	132987
6-lug	64.00	31.60	48.922	137967
7-lug	64.19	32.58	49.922	139340
8-lug	67.00	31.70	52.123	135387
9-lug	72.14	36.50	54.307	143539
10-lug	57.16	27.96	46.866	134816
<b>Media</b>	<b>63.01</b>	<b>29.06</b>	<b>48.02</b>	<b>139887</b>
diff%	-3.36%	6.02%	1.06%	2.21%
diff m/m %	18.51%	172.32%	31.66%	13.83%
diff a/a %	12.66%	74.04%	30.38%	7.70%



### Commento

Al Powernext, nella settimana al 10 luglio, le quotazioni dell'energia hanno segnato una variazione positiva, toccando quota 48,02 €/MWh grazie al rialzo dell'1,06%. In aumento anche la media settimanale dei volumi scambiati, che varia del 2,21% raggiungendo quota 139.887 MWh. Rispetto allo stesso periodo del mese precedente, inoltre, il prezzo medio e i volumi scambiati di energia sono risultati in aumento del 31,66% e 13,83% rispettivamente. Analogamente la variazione tendenziale, che misura la differenza rispetto allo stesso periodo dell'anno prima, è risultata positiva, riportando un +30,38% per il prezzo dell'energia e +7,7% per la quantità di volumi scambiati.

## **3.2 AGENDA GEOECONOMICA**

Al monitoraggio delle borse elettriche si aggiunge la stesura settimanale dell'Agenda Geoeconomica, che consiste nel produrre un documento con i principali dati macroeconomici che usciranno la settimana successiva, riportando il dato passato e le aspettative per quello successivo dei principali dati macroeconomici, indicatimi dallo staff di Axia. L'importanza della conoscenza dei dati macroeconomici serve per avere un ulteriore elemento di conoscenza e monitoraggio del mercato borsistico, che reagisce appunto agli stimoli macroeconomici.

### **3.2.1 ESEMPIO**

Dopo aver ricercato nella rete articoli sull'argomento negli specifici siti di finanza, ecco quanto prodotto per la settimana da lunedì 31 maggio a venerdì 4 giugno 2010:

#### Lunedì 31 Maggio 2010

##### CPI (Euro)

L'inflazione su base annua della zona Euro ad aprile è risultata pari all'1,5%, in linea con le previsioni degli analisti, e in aumento rispetto sia al dato di marzo (+1,4%) che a quello dell'aprile dell'anno scorso (+0,5%). Nell'Europa dei 27 il dato sull'inflazione è del 2%, in crescita rispetto all'1,9% di marzo e al dato di un anno fa di 1,3%. Su base mensile si registra un aumento dei prezzi al consumo dello 0,5% nell'UE-16 e dello 0,4% nell'UE-27.

Considerando i dati dei singoli paesi, l'indice dei prezzi al consumo su base annua ha registrato i tassi più deboli in Lettonia (-2,8%), Irlanda (-2,5%) e Lituania (0,2%), mentre le variazioni più elevate si sono verificate in Ungheria (+5,7%), Grecia (+4,7%) e Romania (+4,2%).

Analizzando i diversi settori i trasporti hanno mostrato l'aumento più elevato, del 5,9%, a causa dei prezzi del carburante, mentre il tasso annuale più debole lo ha fatto registrare il settore dell'intrattenimento e della cultura con il -1%.

Per quanto riguarda le stime future, gli analisti interpellati dalla BCE si attendono un'inflazione nell'Eurozona dell'1,4% nel 2010 e dell'1,5% nel 2011. Per la prima volta è stata data anche una stima per il 2012, con le aspettative che si collocano su un valore medio di 1,7%.

#### Indice di fiducia economica (EUR)

Gli indici di fiducia della zona Euro hanno registrato un rialzo oltre le attese nel mese di aprile. L'indice di fiducia economica ha raggiunto quota 100,6 punti dai 97,9 del mese precedente, superando in consensus che si attestava a 99,4 punti.

Nello specifico, la fiducia dei consumatori ad aprile è salita da -17 a -15 punti, secondo le attese. L'indice di fiducia delle imprese è invece andato oltre le previsioni di -8 punti, raggiungendo quota -7, dopo il -10 di marzo. Anche l'indice di fiducia del settore dei servizi ha superato il consensus degli analisti, che lo davano a 3 punti, attestandosi a quota 5 punti dall'1 di marzo. Sale anche l'indice di fiducia sintetico a +0,23 dopo il dato negativo di marzo di -0,20 (dato rivisto), ben oltre la previsione di -0,12 punti fatta dagli analisti. Resta infine invariata la fiducia dei consumatori nel settore delle costruzioni, stabile a -25.

#### Martedì 1 Giugno 2010

##### Tasso di disoccupazione (GER)

L'Agenzia federale per il lavoro di Norimberga ha comunicato che il tasso di disoccupazione della Germania è sceso ad aprile al 7,8%. Il dato smentisce gli analisti che non si aspettavano variazioni rispetto all'8% del mese precedente, e fa registrare la maggior diminuzione dal 2008. Il numero di disoccupati, in termini destagionalizzati, è diminuito di 68.000 unità rispetto alle 42.000 del mese precedente, e ben oltre le 10.000 attese. Secondo gli analisti il calo è dovuto soprattutto al settore del turismo e delle costruzioni, e il presidente della Bundesbank Axel Weber si è dichiarato fiducioso per la ripresa economica tedesca.

##### Indice ISM manifatturiero (USA)

L'ISM, l'indice sull'attività manifatturiera americana, è cresciuto in aprile a 60,4 punti dai 59,6 (rivisti) del mese di marzo. Il dato, che risulta essere il migliore dal giugno 2004, è andato oltre le attese degli analisti, che si aspettavano una crescita a 60,1 punti. Le varie componenti dell'indice sono aumentate rispetto al mese precedente: i nuovi ordini da 61,5 a 65,7, la produzione da 61,1 a 66,9 e l'occupazione da 55,1 a 58,5 punti. Unico dato in calo il sottoindice delle scorte, che passa a 49,4 punti rispetto ai 55,3 del mese precedente.

Venerdì 4 giugno 2010

### Tasso di disoccupazione (USA)

Sale il tasso di disoccupazione negli USA ad aprile al 9,9%, rispetto al 9,7% del mese precedente (che era anche la previsione degli analisti), e crescono di 290 mila unità i posti di lavoro non agricoli, record mensile dal marzo 2006. Sembra una contraddizione, ma il Dipartimento del Lavoro americano spiega che l'aumento percentuale è dovuto al reinserimento nel mercato del lavoro da parte dei cosiddetti "scoraggiati", ovvero tutti quei lavoratori che avevano rinunciato a cercare lavoro.

Il numero di nuovi posti di lavoro creati va oltre le attese degli analisti (180 mila unità), ma ancora più importante è il dato che 230.000 di questi sono giunti dal settore privato. Il Dipartimento del Lavoro Usa ha anche rivisto al rialzo il numero dei nuovi occupati di marzo portandolo a 230.000 da 162.000 della prima rilevazione.

### **3.3 PETROLIO E COMMODITIES**

Al commento dei dati macroeconomici si aggiunge il monitoraggio dell'andamento del mercato delle commodities e del petrolio. Il termine commodity indica quei beni per cui vi è un'offerta sul mercato senza differenze qualitative, e che abbia le stesse caratteristiche indipendentemente da chi le produce. Un'altra peculiarità importante è che deve essere stoccabile e conservabile nel tempo, e che non vi sia una lavorazione elaborata per ottenere il prodotto. Una delle proprietà fondamentali delle commodities è che il mercato ne determina il prezzo. Le principali commodity negoziabili sui mercati internazionali possono essere suddivise in: agricoli (avena, soia, mais, ecc.), coloniali e tropicali (cacao, caffè, zucchero), metalli (alluminio, oro, nickel, argento, rame, ecc.), energetici (petrolio, gas naturale, ecc.) e carni (bovini e maiali). Di solito le commodity sono negoziate, tramite future (contratti a termine standardizzati per la Borsa), nel mercato di New York (NYMEX), Chicago (CBOT), Londra (LME), ecc.

In particolare il mio compito era quello di trovare informazioni sull'oro e su quei beni che avevano un andamento significativo (sia in positivo che in negativo), oltre che il monitoraggio delle scorte di petrolio Usa e il suo andamento sul mercato, in quanto si tratta di una delle commodity più importante capace di determinare con il suo prezzo l'andamento delle Borse in tutto il mondo.

### **3.3.1 ESEMPIO: COMMENTO**

In queste settimane l'oro ha attratto gli investitori di tutti il mondo per la sua nota caratteristica di bene rifugio. La settimana appena trascorsa non ha fatto eccezioni: dopo i valori minimi fatti registrare venerdì scorso (1166,3\$ per oncia) a causa della volontà degli investitori di ripararsi dalle oscillazioni del mercato, il prezzo del metallo prezioso è risalito a quota 1212\$/oncia giovedì mattina, dopo i massimi fatti registrare mercoledì.

La risalita del metallo prezioso è dovuta a diversi fattori. Tra questi c'è senz'altro l'annuncio dell'Italia di un piano di riduzione del budget, che gli analisti ipotizzano possa creare difficoltà nella ripresa domestica e aumentare ulteriormente i problemi dell'Eurozona, e che ha causato la discesa dell'euro. In aggiunta c'è da registrare il calo oltre le aspettative della fiducia dei consumatori tedeschi, la nazionalizzazione delle banche di risparmio CajaSur in Spagna e le sempre più alta tensione tra le due coree.

A questi fattori va aggiunto che il World Gold Council ha previsto una forte crescita della domanda dell'oro dopo il calo dell'11% registrato nei primi tre mesi del 2010. Anche i gioielli sembrano tornare di moda: nel primo trimestre di quest'anno il settore ha consumato il 43% in più di tonnellate d'oro, per un ammontare di 470,7: a trainare il consumo è soprattutto l'India, con una crescita del 291%, seguita da Medio Oriente (+23%) e Cina (+11%). Solo l'offerta di oro da riciclare è diminuita (-43%) compensata solo in parte dalla crescita dell'offerta mineraria (+5%).

Infine da registrare il record di 1267,3 tonnellate di possedimenti auriferi del fondo di scambio più grande al mondo, l'SPDR Gold Trust di New York, che nella sola giornata di mercoledì sono aumentati di 30,4 tonnellate, record che non si registrava dal febbraio 2009.

Parlando anche dei metalli non ferrosi si registra un costante e graduale recupero: soprattutto il rame al London Metal Exchange ha chiuso giovedì con un rialzo dell'11,5%. Al Comex il palladio è cresciuto di 3,6 punti percentuali.

Per le commodities agricole si registra, sia pure di poco, il record del cacao da 32 anni nel mercato di Londra, mentre è salito del 4,2% a New York. Male invece lo zucchero, poiché ci si aspetta un'abbondante raccolta di canne da zucchero in Brasile, così che il grezzo ha perso il 2,9% e quello raffinato il 2,7%. Guadagno quasi dell'1% per caffè robusta e arabica, come anche i semi di soia (+1,5% al Cbot) e il frumento (+1,3%)

In netto rialzo il prezzo del greggio, che tocca il suo massimo da due settimane a questa parte, anche grazie all'attenuarsi delle preoccupazioni sulle sorti dell'Europa: da quota 68,10 \$ al barile del 25 maggio, è arrivato a toccare quota 75,40 dollari al barile nella notte di giovedì, per poi attestarsi ad una quota di circa un dollaro in meno.

La crescita del prezzo del petrolio è motivata anche dall'aumento dei consumi dei prodotti petroliferi negli USA del 6,9%, in particolare del 15,8 % per i distillati: incremento quasi certamente dovuto al trasporto merci, che quindi fa ben sperare per la ripresa dell'economia americana

In aumento le scorte di petrolio negli USA al 22 maggio: lo rende noto il Dipartimento dell'Energia americano, comunicando che le scorte sono aumentate di 2,46 milioni di barili a 365,146 milioni, ben oltre le aspettative degli analisti che si attendevano un aumento di 100.000 unità dopo il rialzo di 200.000 della settimana precedente. Gli stock della benzina sono invece diminuiti di 203.000 barili, attestandosi a quota 221,631 milioni, proseguendo il ribasso registrato sette giorni prima di 300.000 barili. Nonostante le stime fossero per un incremento di 300.000 unità dopo il calo di 1 mln registrato precedentemente, anche le scorte di distillati sono scese di 267.000 barili a 152,543 milioni. Infine l'utilizzo della capacità degli impianti è sceso di circa 0,1 punto percentuale, da 87,94% a 87,8%. Non ci sono state infine variazioni per quanto riguarda le riserve strategiche.

## **4. APPROFONDIMENTO: SYNTHEMA 2.0**

### **4.1 COS'E' SYNTHEMA 2.0**

Lo scopo dell'utilizzo della piattaforma Synthema è quello di poter disporre delle informazioni non strutturate presenti nel web, come articoli di giornale o agenzie di stampa, per l'analisi, estrazione e classificazione delle informazioni in essa presenti. Senza tale supporto è possibile un accesso semplice ed intuitivo alle informazioni solo se organizzate in dati, memorizzati in tabelle o database. Dai motori di ricerca, inoltre, il recupero delle informazioni non è semplice, e spesso restituisce risposte alle nostre interrogazioni di cui non ne è chiara né la rilevanza né la pertinenza. L'analisi linguistica conducibile tramite Synthema permette dunque di cogliere i concetti chiave di un testo, svolgendo elaborazioni sintattiche, morfologiche, logico funzionali e semantiche. L'elaborazione morfo-sintattica permette di risolvere le eventuali ambiguità presenti nel testo. Con l'analisi logica si identifica il ruolo svolto nel testo da ciascuna parola, permettendo di capire ogni volta chi fa che cosa, dove, come e quando. Infine l'aspetto semantico dell'analisi, che consente di cogliere il significato di ogni parola in maniera più approfondita.

Di seguito all'analisi linguistica viene fatta un'analisi statistica, che permette di assegnare i documenti "concettualizzati" tramite Clustering a categorie tematiche non predefinite e tramite Categorizzazione a categorie personalizzate e definite a priori. Vi è inoltre la possibilità di osservare la distribuzione temporale dei documenti e approfondire i legami tra le varie parole tramite grafici di prossimità e relazione.

### **4.2 COME FUNZIONA SYNTHEMA 2.0**

Nelle prossime righe illustrerò brevemente come si procede per un'analisi tipo con Synthema, per illustrare quali siano le fasi principali dell'analisi morfo-sintattica, le sue funzioni e strumenti principali.

#### **4.2.1 FASE 1: RACCOLTA DATI**

La prima fase è quella dell'inserimento dei siti di interesse. La piattaforma è dotata di una schermata attraverso la quale è possibile istruire il programma sui siti dai quali prendere informazioni. In aggiunta si può anche richiedere che i commenti inseriti dai vari utenti vengano suddivisi, in modo che sia possibile in seguito distinguere un commento positivo da uno neutro o

negativo. Il programma procede quindi con la scansione del sito stesso o Crawling, e analizza il testo presente dal punto di vista lessicale e semantico, nonché contando le ricorrenze delle parole e datando gli articoli e i commenti.

#### **4.2.2 FASE 2: BASE DI CONOSCENZA**

Fatto questo si può procedere con la seconda fase, che consiste nell'affinare la base di conoscenza: si correggono quindi eventuali errori del sistema e si aggiungono le relazioni concettuali che si reputano mancanti; se necessario, infine, si definiscono le parole ambigue che possono avere più significati.

#### **4.2.3 FASE 3: ANALISI DEI DATI**

La terza fase consiste nell'analisi dei dati, utilizzando gli strumenti che Synthema mette a disposizione, di cui i principali sono: funzioni di interrogazione e ricerca nei risultati della scansione, clustering, distribuzione temporale dei documenti e distribuzione statistica per concetti e relazioni. E' anche possibile la consultazione dei documenti, attraverso la quale si può chiarire ulteriormente il contesto e la connotazione delle parole.

### **4.3 ESEMPIO: COMMENTI DEGLI UTENTI DELLA RETE SUL RISULTATO SPORTIVO DELLA CHAMPIONS LEAGUE 2009/2010.**

Di seguito verrà riportato un esempio di utilizzo della piattaforma Synthema 2.0. L'argomento scelto è la Champions League 2009/2010, con particolare attenzione ai pareri degli utenti della rete riguardo la vittoria dell'Inter nella finale di Madrid del 22 maggio 2010.

#### **4.3.1 RACCOLTA DATI**

Per prima cosa ho inserito nel programma i siti internet dai quali scaricare le informazioni. Ho scelto di prestare particolare attenzione ai blog e ai forum della rete, in quanto riportano al meglio le opinioni dei tifosi: un articolo di giornale, infatti, è un testo che relaziona su un avvenimento, presenta dei fatti o espone un punto di vista, ma di una sola persona: il giornalista. Al contrario, nei blogs e nei forum, o nei siti dove è possibile commentare l'articolo di giornale, gli utenti possono interagire tra loro, potendo esprimere un'opinione a riguardo. La scelta dei siti da inserire è stata

fatta utilizzando la ricerca tramite <http://www.google.it> e <http://blogsearch.google.com>. Le parole chiave utilizzate sono state “Champions League” o “Finale Madrid”. Nella ricerca ho deciso di porre anche altri vincoli, per non ottenere un numero troppo elevato di risultati difficile da analizzare nei momenti successivi. Il periodo considerato, per esempio, va dal 15 maggio 2010, una settimana prima della finale, al 10 luglio 2010. Ho inoltre limitato la ricerca ai soli siti in lingua italiana. Infine ho prestato attenzione ai siti che caricavo, escludendo quelli non pertinenti con l’argomento scelto.

Per la suddivisione dei commenti, nel momento dell’inserimento, era necessario descrivere al programma, in linguaggio html, quale fosse il codice per la divisione: per farlo è stato necessario utilizzare il Browser Mozilla, ed il programma Firebug, che consente di visualizzare in parallelo alla pagina il codice Html.

Per motivi di dimensione non è stato caricato l’intero sito, ma la pagina o le pagine che contenevano i commenti e gli articoli da prendere in considerazione. Di seguito vengono riportati i siti dai quali sono stati tratti gli articoli e i commenti:

Dopo aver caricato un sito, la piattaforma Synthema procede con l’analisi dei contenuti, e si crea la base di conoscenza che servirà poi per lo studio dell’argomento prescelto. In questa fase è molto importante controllare che gli articoli e i commenti vengano caricati correttamente, e non, per esempio, con la data sbagliata o includendo parti del sito che risultino dannose per la ricerca. Nel caso specifico ho dovuto prestare attenzione alla datazione e divisione dei commenti, e a controllare che venissero caricate le pagine indicate e non anche altri articoli raggiungibili tramite link presenti nella pagina.

#### **4.3.2 PERFEZIONAMENTO DELLA BASE DI CONOSCENZA**

La seconda fase, come illustrato sopra, serve ad affinare la base di conoscenza. Nel mio caso, per esempio, è stato necessario istruire il sistema che la parola “Champions League” era equivalente a “Uefa Champions League”, “Champions” o “Coppa campioni”, come si vede nella figura 1.

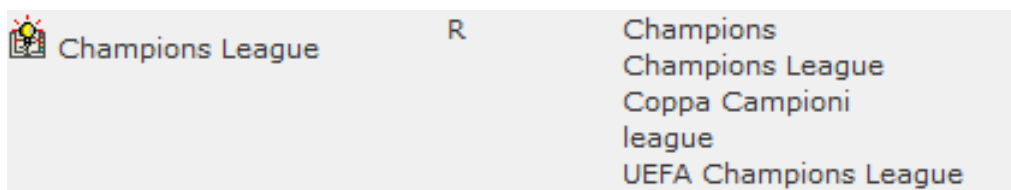


Figura 1: perfezionamento della base di conoscenza



Inoltre, essendo che il concetto può essere espresso in diversi contesti, si rende necessario aggiungere alcune relazioni se non sono già presenti, come quelle di iponimia (quando il significato della seconda parola, più generale, è incluso nel significato della prima, più specifica) o di traduzione, come riportato dall'esempio nella figura 2.



Figura 2: relazioni con il termine Champions League

*Nota : la percezione dell'ironia presente in alcuni commenti degli utenti non viene rilevata da Synthema, che in certi casi appunto non è in grado di comprendere appieno il significato della frase. Sta all'utente quindi, quando è possibile, impartire al sistema le istruzioni per dissolvere ogni ambiguità.*

### **4.3.3 ELABORAZIONE ED ANALISI DEI DATI**

Creata dunque una base di conoscenza, si può procedere ora con la vera e propria analisi, utilizzando gli strumenti di Synthema prima elencati. La mia ricerca si è focalizzata sulla percezione della rete riguardo la finale di Champions League. Il mio intento è quello di capire come i tifosi delle squadre di calcio abbiano reagito alla vittoria dell'Inter in finale: se prevale una reazione di gioia per il trionfo di una squadra italiana (che consente tra l'altro al Campionato Italiano di mantenere 4 posti per la Champions League del 2011/2012) o se è maggiore l'invidia e l'amarezza per la vittoria di una squadra rivale, o ancora se la reazione risulta essere neutra.

#### **4.3.3.1 GRAFICI RELAZIONALI**

Iniziamo con l'analisi tramite grafo relazionale: il programma consente infatti di visualizzare le parole (colorate a seconda che siano nomi, verbi, avverbi, ecc.) che sono in rapporto con il termine

di interesse con una determinata frequenza; dal colore delle frecce è inoltre possibile conoscere il tipo di legame concettuale (per esempio: A viene qualificato da B, A si svolge nel luogo B, ecc.). Qui di seguito (figura 3) è riportato il grafo relazionale che ha come termine centrale il nome proprio Champions League. Vi sono vari termini legati all'evento, tutti con frequenza minima 4, cioè la relazione si presenta almeno 4 volte nei documenti che hanno al loro interno la parola chiave della ricerca.



Figura 3: grafico relazionale: parola chiave “Champions League” – frequenza minima 4

Tra le paole legate al termine Champions League c'è anche la parola “vincere”. Se clicco sopra quest'ultima il grafico (figura 4) si arricchisce con le relazioni che sono ad essa legate, sempre con frequenza minima 4.

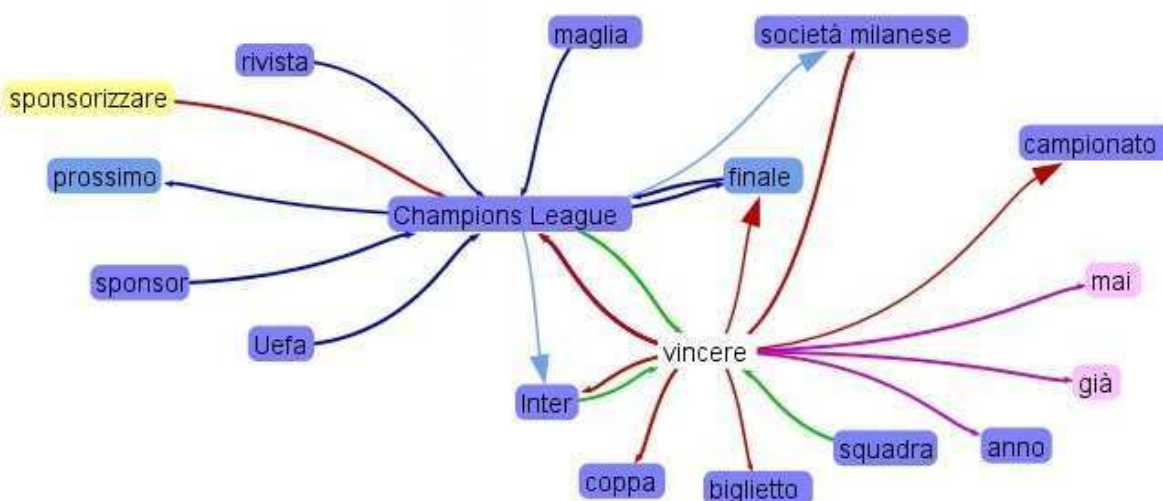


Figura 4: grafico relazionale: “espansione” della parola “vincere” – frequenza minima 4

Vediamo come i due lemmi siano entrambi legati alla parola “Inter”, che infatti è risultata essere la squadra *vincitrice* della *Champions League* nella *finale* di Madrid. Questo ci permette già di affermare che la ricerca ci consentirà di stabilire il sentiment della rete, in quanto partendo da parole chiave generali quali “Champions League” o “Finale Madrid” è emerso anche l’oggetto principale della nostra ricerca: l’Inter.

Per approfondire meglio questo aspetto visualizziamo il grafico (figura 5) con frequenza minima 4 la cui parola centrale è “Inter”: non emergono lemmi particolarmente significativi per la ricerca, se non il concetto di “beneficiare” che vedremo più avanti.

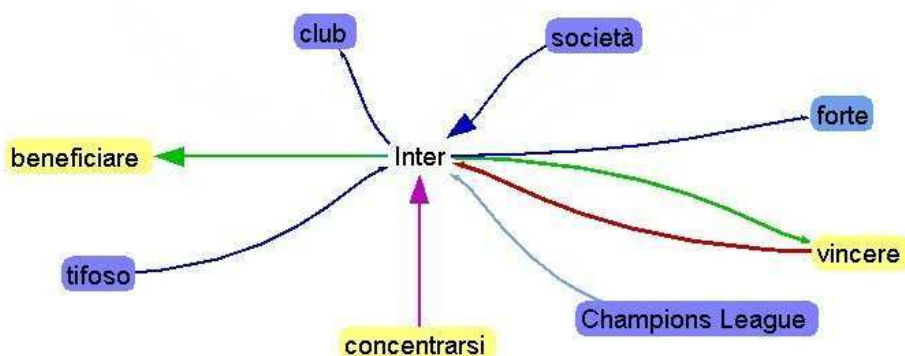


Figura 5: grafo relazionale: parola chiave “Inter” – frequenza minima 4

Passiamo ora ad un termine che è particolarmente interessante per capire la percezione dei tifosi della rete riguardo tale successo. Cercando tra i documenti la parola chiave “complimento” otteniamo il seguente grafico (figura 6), con frequenza minima però ridotta a 1.



Figura 6: grafo relazionale: parola chiave “complimento” – frequenza minima 1

Se selezioniamo le frecce che mettono in relazione semantica le parole tra loro, il programma ci consente di visualizzare la frase che le lega: in questo caso selezioniamo nell'ordine le frecce che collegano “complimento” con “dire”, “mandare” e “raggiungere” (figura 7)

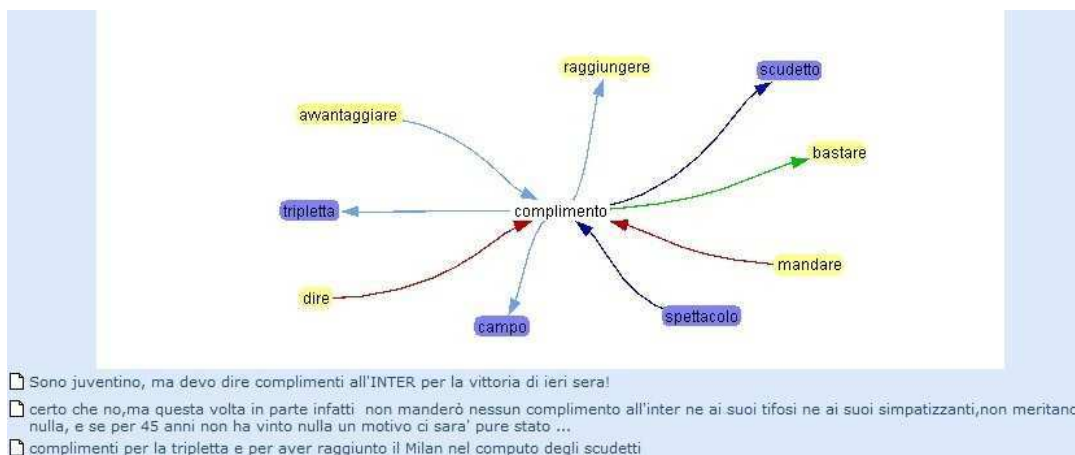


Figura 7: visualizzazione delle frasi che contengono la relazione tra due diversi lemmi

L'icona con il foglio a inizio citazione serve per consultare il documento dal quale è stata estratta la frase, come si può vedere per il primo commento dall'immagine di figura 8.



Figura 8: visualizzazione del sito dal quale proviene il commento che ha dato origine alla relazione

Altri grafici a cui è interessante prestare attenzione sono quelli che parlano dell'arbitraggio, uno dei maggiori argomenti di discussione nei forum calcistici e non solo. Se poniamo come concetto centrale appunto l'arbitraggio, otteniamo il seguente grafico (figura 9):

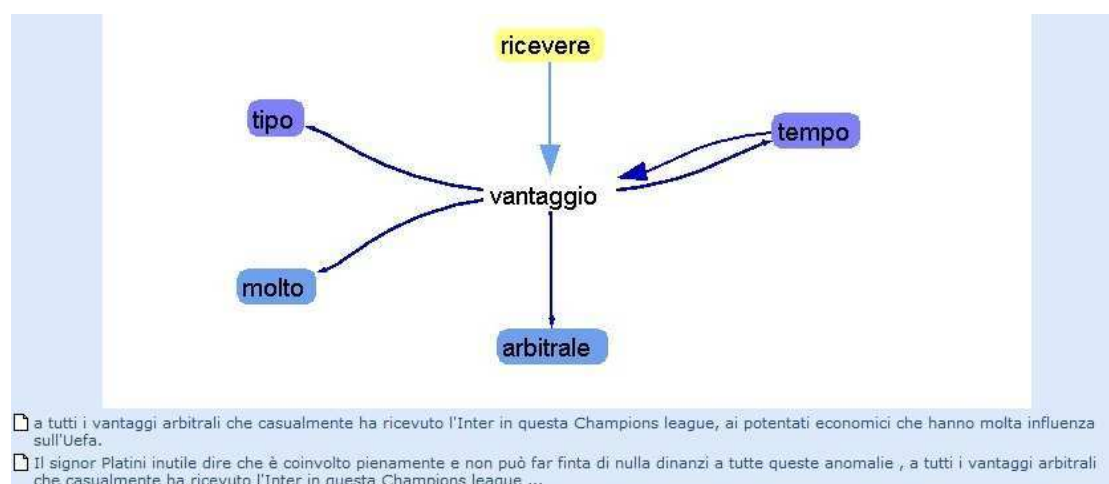


Figura 9: grafo relazionale: parola chiave “vantaggio” – frequenza minima<sup>3</sup>

In generale questo modo di procedere consente di capire l’accezione positiva, negativa o neutra delle parole, ma non ci consente di quantificarle e confrontarle tra loro, o di collocarle nel tempo, in quanto dovremmo cercare per ogni termine che riteniamo significativo il relativo grafico, e per ognuno di essi studiare le varie relazioni presenti e i relativi documenti nei quali sono presenti.

#### **4.3.3.2 DISTRIBUZIONE TEMPORALE DEI PRINCIPALI TERMINI DI INTERESSE**

Per procedere dunque nell’analisi si rendono utili altri strumenti che Synthema mette a disposizione. Il primo di questi è la distribuzione temporale dei termini. Quando la piattaforma scarica un documento, infatti, questo viene datato secondo la data di pubblicazione: è quindi possibile vedere l’evoluzione nel tempo del termine considerato, confrontandolo anche con altri concetti. Successivamente è possibile ottenere un grafico a torta relativo ai concetti considerati.

Per l’esempio di lavoro ho pensato che fosse rilevante conoscere la distribuzione temporale dei termini che hanno accezione positiva o negativa, e che quindi essendo estrapolati da documenti riguardanti la “Champions League” o “Finale Madrid” (che dal grafo relazionale di figura 4 abbiamo visto essere collegata all’Inter) possano darci indicazioni sulla percezione della finale e dei vincitori. La distribuzione temporale non è altro che il numero di documenti nei quali compare il lemma che ci interessa, indicizzato con la data del documento, e non il numero di ricorrenze con il quale il termine viene ripetuto nei documenti. Il grafico che ne risulta è riportato nella figura 10.

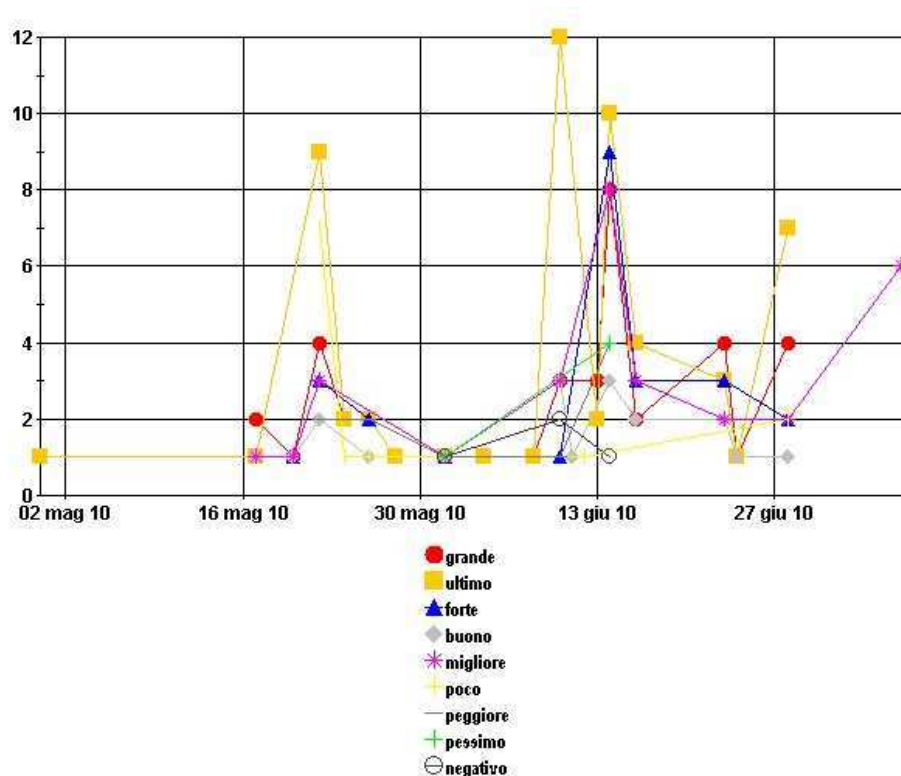


Figura 10: distribuzione temporale degli aggettivi “positivi” e “negativi”

La distribuzione temporale mostra innanzi tutto che l’andamento degli aggettivi non differisce tra quelli positivi e negativi. Entrambi mostrano due picchi, uno in prossimità o nel giorno della finale del 22 maggio 2010, e l’altro attorno all’inizio della terza settimana di giugno, tra il 13 e il 15 principalmente: il secondo picco è dovuto all’inizio del Campionato del Mondo di calcio, principale evento calcistico che ha seguito la Champions League, e alla conseguente ripresa delle discussioni in merito calcistico dopo la finale di Madrid.

#### **4.3.3.3 GRAFICI A TORTA PER I TERMINI DI INTERESSE**

Dalla distribuzione temporale non è facile però capire quale termine prevalga, e se in generale vi sia maggioranza di termini positivi o negativi. Per avere quindi un’idea più chiara consideriamo i termini che stiamo prendendo in considerazione nell’arco di tempo stabilito per la ricerca, e, sempre grazie a Synthema, possiamo rappresentarli e confrontarli in un grafico a torta (figura 11). Le percentuali sono calcolate in base al numero di documenti nel quale compare il lemma sul totale dei documenti nei quali è presente almeno uno dei termini presi in considerazione.

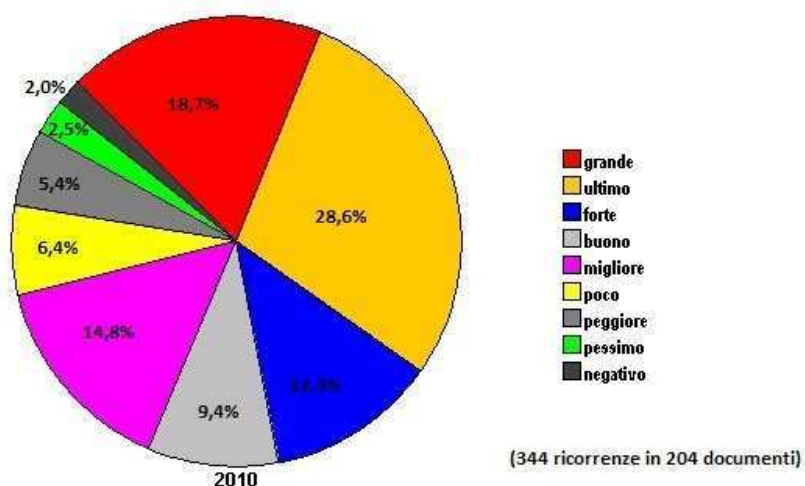


Figura 11: grafico a torta degli aggettivi “positivi” e “negativi”, con relative percentuali

Nella tabella 1 ho suddiviso gli aggettivi nelle due categorie. Vediamo come la maggioranza sia nettamente a favore di quelli con accezione positiva.

AGGETTIVO “POSITIVO”	N° Doc.	Percentuale	AGGETTIVO “NEGATIVO”	N° Doc.	Percentuale
Grande	38	18,7	Ultimo	58	28,6
Forte	25	12,3	Poco	13	6,4
Buono	19	9,4	Peggior	11	5,4
	30	14,8	Pessimo	5	2,5
			Negativo	4	2,0
<b>TOTALE*</b>	<b>112</b>	<b>55,2</b>		<b>91</b>	<b>44,9</b>

Tabella 1: suddivisione e confronto tra gli aggettivi “positivi” e quelli “negativi”.

\* Il totale può non dare 100 a causa di approssimazioni

Un ulteriore conferma che gli aggettivi siano da attribuire alle squadra di Mourinho è la prevalenza che ha quest’ultima rispetto alle principali altre squadre nelle ricorrenze dei documenti che risultano dalla ricerca con parola chiave “Champions League”. Come si può vedere (tabella 2) solo la somma di almeno 7 squadre è in grado di superare la percentuale dell’Inter, che è del 42,7% nei 108 documenti su 250 trovati (massimo di risultati impostato per la ricerca), con ben 288 ricorrenze (vedi figura 12).

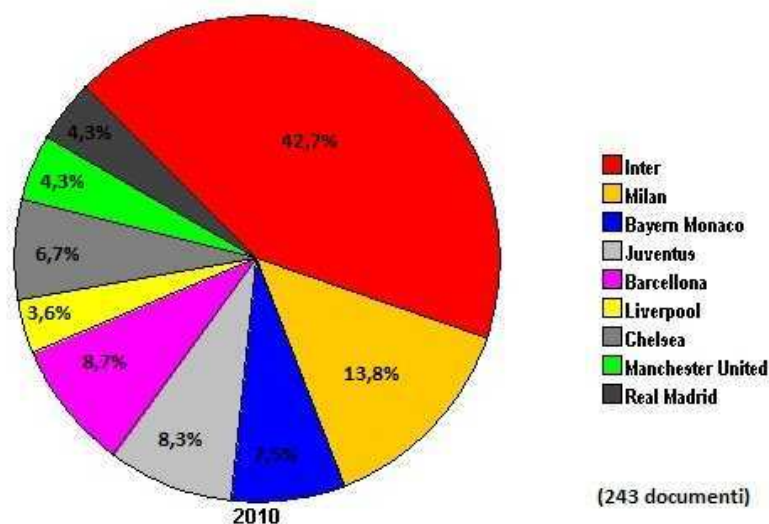


Figura 12: grafico a torta relativo alle principali compagini calcistiche europee

	N° Doc.	Percentuale			N° Doc.	Percentuale
<i>INTER</i>	108	42,7	<i>ALTRE SQUADRE</i>		145	57,2
				Milan	35	13,8
				Bayern Monaco	19	7,5
				Juventus	21	8,3
				Barcellona	22	8,7
				Liverpool	9	3,6
				Chelsea	17	6,7
				Manchester Utd	11	4,3
				Real Madrid	11	4,3

#### **4.3.4. CONCLUSIONI**

Nelle pagine precedenti si è visto come la ricerca con parole chiave “Champions League” e “Finale Madrid” in determinati blog e forum in lingua italiana abbia prodotto dei risultati che mostrano come l’Internazionale sia uno degli argomenti centrali della discussione. Abbiamo infatti visto come sia presente come unica squadra nella relazione con la parola “Champions League” e come nei documenti sia quella più volte citata e presente nella maggior parte di essi (quasi la metà, con il 42,7%). Se quindi possiamo affermare che in più del 40% dei documenti l’ex squadra di Mourinho viene discussa, è altamente probabile che gli aggettivi siano principalmente ad essa riferiti. Essendo questi in maggioranza positivi è dunque possibile affermare che la reazione degli utenti della rete



alla vittoria dei nerazzurri è positiva. Il risultato è bilanciato dalla gioia e l'euforia dei tifosi interisti unita ai complimenti degli utenti più "sportivi" con l'amarezza e l'invidia dei tifosi delle altre squadre, che evidentemente non gradiscono tale successo; non si registra, infine, un numero di commenti o aggettivi neutrali di particolare rilevanza.

## **5. SITOGRAFIA**

[http://it.wikipedia.org/wiki/Pagina\\_principale](http://it.wikipedia.org/wiki/Pagina_principale)  
<http://www.google.it>  
<http://www.passionemaglie.it>  
<http://www.fashiontimes.it>  
<http://gilveus.blogspot.com>  
<http://sportevai.it>  
<http://www.calcio mercatonews.com>  
<http://www.calciami.it>  
<http://www.asromalive.com>  
<http://www.sportstadio.com>  
<http://nuovoindiscreto.blogspot.com>  
<http://jobtalk.blog.ilsole24ore.com>  
<http://www.climarsoft.com/forum>  
<http://www.corederoma.it>  
<http://screwdrivers.splinder.com>  
<http://criticarossonera.blogspot.com>  
<http://newsfromtshirts.wordpress.com>  
<http://www.youmark.it>  
<http://dolcefarnuoto.it>  
<http://www.dgmag.it>  
<http://www.blogseriea.it>  
<http://rettilineotribuna.it>  
<http://feeds.blogosfere.it>  
<http://abasto.splinder.com>  
<http://www.megacartedicredito.com>  
<http://nuovoindiscreto.blogspot.com>  
<http://schirone.blogspot.com>  
<http://www.blogseriea.it>  
<http://feeds.blogoi.it>  
<http://www.calcio-oggi.it>  
<http://febbrea90magazine.blogspot.com>  
<http://nuovoindiscreto.blogspot.com>  
<http://www.blogseriea.it>  
<http://www.newnotizie.it>  
<http://criticarossonera.blogspot.com>  
<http://www.blogseriea.it>  
<http://www.altrosport.com>  
<http://www.romacalcionews.it>  
<http://www.oleole.it>  
<http://www.dimmicosacerchi.it>  
<http://auladuecentosedici.blogspot.com>  
<http://www.ciaojuve.com>  
<http://www.oasisblues.com>  
<http://gold.libero.it>  
<http://blog.libero.it>  
<http://www.forzaroma.info>  
<http://zazza.blog.deejay.it>  
<http://blog.libero.it>  
<http://www.sporth24.com>  
<http://www.forza-roma.com>  
<http://www.estense.com>  
<http://testedicalcio.blogosfere.it>  
<http://blog.ju29ro.com>  
<http://www.bausciacafe.com>  
<http://jucirdanro.blogspot.com>  
<http://azzurrissimonapoli.blogspot.com>  
<http://www.facebook.com>  
<http://www.finanzaonline.com/forum>  
<http://twitter.com>  
<http://www.youtube.com>  
<http://www.myspace.com>