

USING THE WEB TO MONITOR A CUSTOMIZED UNIFIED FINANCIAL PORTFOLIO

Camilo Restrepo-Arango

Arturo Henao-Chaparro

Claudia Jiménez-Guarín

Department of Systems and Computing Engineering

Universidad de los Andes - Bogotá, Colombia

Ninth International Workshop on Web Information Systems Modeling (WISM 2012)

31st International Conference on Conceptual Modeling (ER 2012)

October 15th – 18th, 2012, Florence, Italy

Agenda

- Motivation
- The Problem
- The Context
- Customized Unified Financial Portfolio
 - Web information integration features
 - Data model for customized queries
- The Proposed Solution
- Test and Results
- Conclusions and Future Work

Motivation

- Discovering of relevant and pertinent information on the Web
 - Flooding, discrimination, discovering, gathering
- Domain-specific searching and information integration
 - Financial personalized information
- Financial services and information for several audiences
 - Google Finance
 - Yahoo! Finance
 - Bloomberg
 - ...
- Usual general market trends information
 - Dynamic, factual
 - Usually stocks, news, RSS feeds
 - Ontologies, news analysis
 - Not social and crowdsourcing
- Lack of meaningful information integration



The Problem

- Filtering, gathering, discovering
 - On time
 - When needed
 - What is needed
- Not only financial facts are relevant
 - Analysis publishing, blogs, social media
 - Web 2.0 collaborative knowledge generation
- Personalization
 - User portfolios
 - Pertinent investments
 - More than stocks
- **Perceptions, trends**



The Context

- Use of Web information retrieval approaches
- Content is the more important feature
 - Integrated to structured data
 - Content is really more than tags and titles
- Sources are heterogeneous, autonomous
- No suppositions can be made about
 - Structure
 - Actual content description
 - Language quality
- Multilanguage sources
- Big data and Web 2.0 techniques are useful and appropriate



Customized Unified Financial Portfolio (UFP)

- Mash-up integrating 242 + 4 financial information sources
- Content is the basis for retrieval
- User financial portfolios filtering are considered
- Integration of detailed financial information, social trends and news for each portfolio asset.
- Big Data approach considering sources: V^3
- Scalability, flexibility
- Information freshness and relevance vs. performance
- Content gathering, indexing, storing
- Semantic analysis considering social media sources

Information Integration Features

- 3 Categories of information, in 2 languages
 - Stock exchange sources: BVC, NYSE
 - Newspapers from several countries
 - Portafolio – Colombia
 - The Economist – England
 - New York Times – USA
 - The Wall Street Journal – USA
 - Crawled news feeds
 - Social networks: Facebook, Twitter

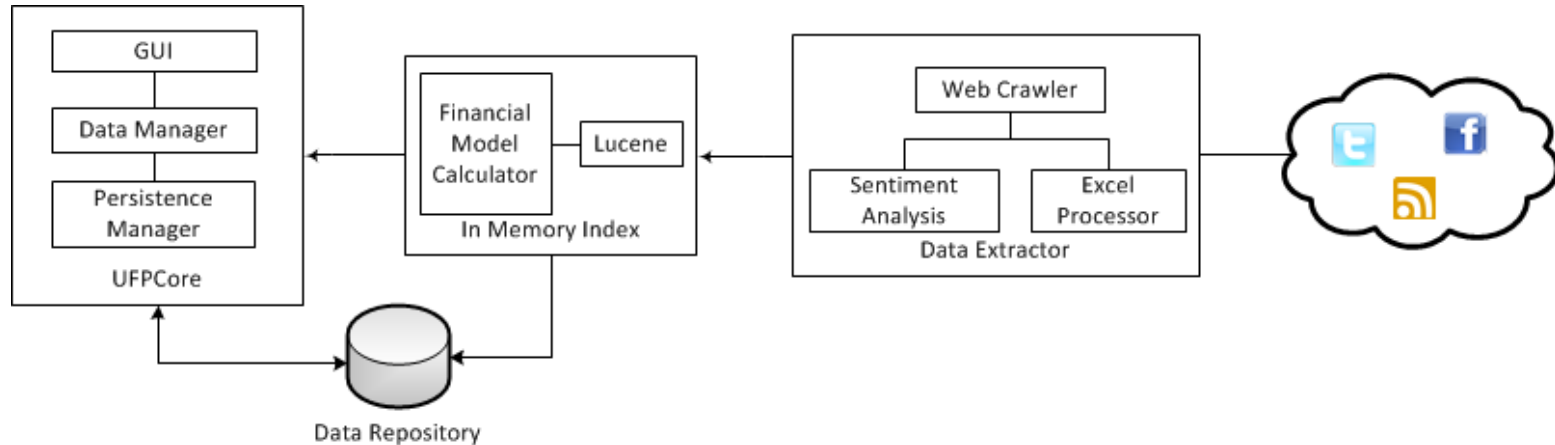
UFP Process

UFP process the information in three stages

1. Configuration
 - A. initial interesting data sources
 - B. Financial model established by a domain expert
2. Source crawling, information classification and indexing
3. Monitor of user relevant information concerning his portfolio

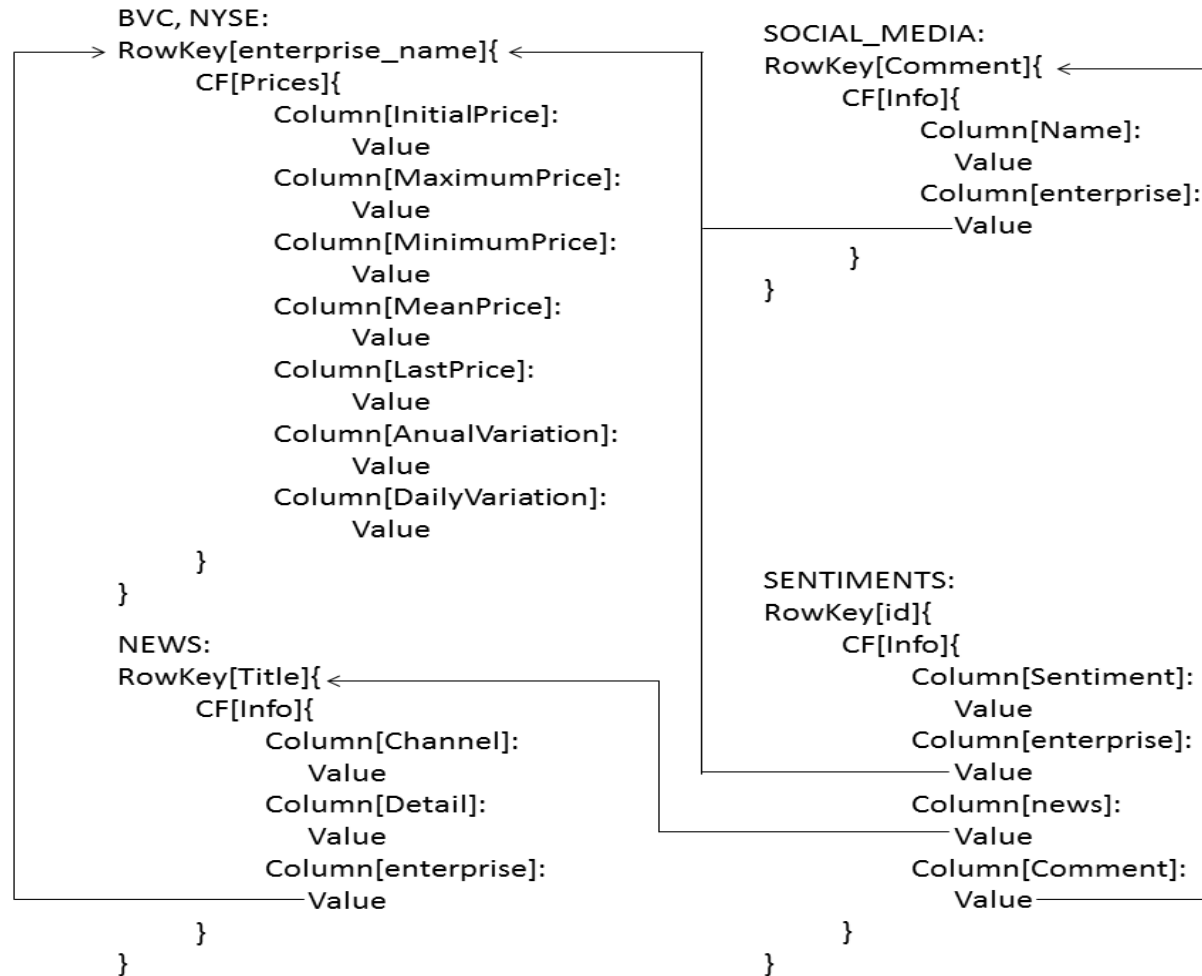
Example: [Skandia Multifund Portfolios](#)

UFP Application Architecture



- UFPCore
- Data Extractor
 - Specialized crawlers and analyzers for each information source type and language
 - Sentiment analysis of social media sources
- Information Retrieval Model
 - Full text search and indexing – Syntactic heterogeneity
 - Enterprise and relevant documents matching
- NoSQL column store repository

Financial Data Model



Asset Semantic Table Definition Process

Financial Expert Definition

Enterprise Asset: Ecopetrol

Petroleum	INDUSTRY	1. Define relevant keywords associated with the industry.	8. Assign weights to every keyword associated with the industry	100%
Exploration Refinery Oil	MAIN SERVICES	2. Define relevant keywords associated with the main services of the company.	7. Assign weights for every keyword associated with the main services of the company	20% 50% 30%
Coil Environment Renewable energy	RIVAL BUSINESS	3. Define relevant keywords associated with the rival business.	6. Assign weights to every keyword associated with the rival business.	40% 20% 40%
EC BVC NYSE	STOCK EXCHANGE	4. Define relevant keywords associated with the stock exchange.	5. Assign weights to every keyword associated with the stock exchange	70% 20% 10%
9. Assign weights to each topic where the keywords where assigned				

Industry: 20% - Main Services: 30% Rival Business: 20% Stock: 30%

Data Model for Customized Queries

Ecopetrol



Topic	Weight 100%	Keywords	Keywords relative weight	Keywords absolute weight
Industry <i>ind</i>	w_{ind} 20%	kw_{ind_1} kw_{ind_2} kw_{ind_n} Petroleum	rkw_{ind_1} rkw_{ind_2} rkw_{ind_n} 100%	tkw_{ind_1} tkw_{ind_2} tkw_{ind_n}
Main Services <i>ms</i>	w_{ms} 30%	kw_{ms_1} kw_{ms_2} kw_{ms_n}	rkw_{ms_1} rkw_{ms_2} rkw_{ms_n} 100%	tkw_{ms_1} tkw_{ms_2} tkw_{ms_n}
Rival Business <i>rb</i>	w_{rb} 20%	kw_{rb_1} kw_{rb_2} kw_{rb_n}	rkw_{ms_1} rkw_{ms_2} rkw_{ms_n} 100%	tkw_{ms_1} tkw_{ms_2} tkw_{ms_n}
Stock Exchange <i>se</i>	w_{se} 30%	kw_{se_1} kw_{se_2} kw_{se_n}	rkw_{ms_1} rkw_{ms_2} rkw_{ms_n} 100%	tkw_{ms_1} tkw_{ms_2} tkw_{ms_n}

Document score in customized queries

$$topic = [w_{topic}, \{kw_{topic_i}, rkw_{topic_i}, tkw_{topic_i}\}, \{kw_{topic_i}, rkw_{topic_i}, tkw_{topic_i}\}] \neq \emptyset \quad (1)$$

$$\sum_{i \in topics} w_i = 1 \quad (2)$$

$$\sum_{i \in KW_t} rkw_{t_i} = 1, \forall t \in topics \quad (3)$$

$$tkw_{t_i} = w_t * rkw_{t_i}, \forall t \in topics, \forall i \in KW_t \quad (4)$$

- When UFP retrieves the information about a specific asset, the semantic table is used to search not only the asset name, but also the associated keywords.
- Weights are used to calculate the final document score given an asset.
- Equations 2, 3, 4 define the relative importance of each topic keyword as well as the importance of an asset topic

Document score in customized queries ...

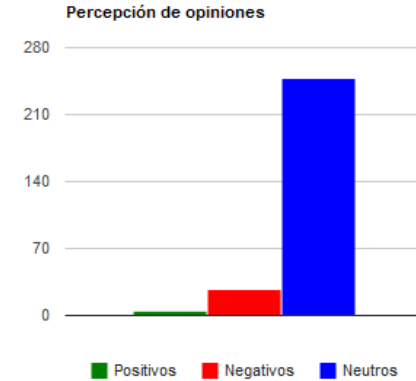
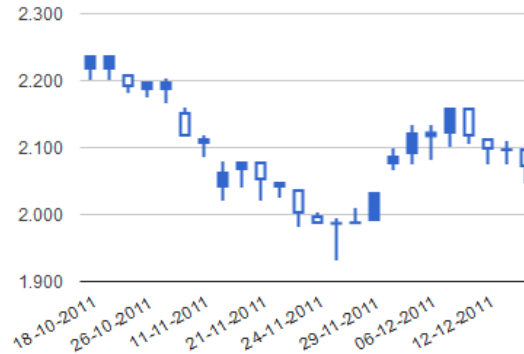
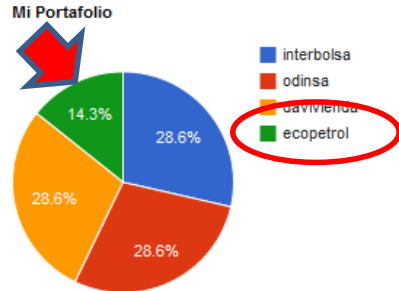
$$UFPscore_{(q,d)} = C_{(q,d)} \cdot norm_{(q)} \cdot sumt_{(q,d)} \quad (5)$$

$$norm_q = \frac{1}{\sqrt{\sum_{t \in q} (idf_{(t)} \cdot tkw_{(t)})^2}} \quad (6)$$

$$sumt_{(q,d)} = \sum_{t \in q} (tf_{(t,d)} \cdot idf_{(t)}^2 \cdot tkw_{(t)}) \quad (7)$$

- The Vector-Space Model $C_{(q,d)}$: how many keywords are found in the document
- $norm_{(q)}$: normalizing factor making comparable scores between queries, using the inverse document frequency sum for all the terms in order to convert the final score to a normal form
- $sumt_{(q,d)}$: contribution of each keyword to the query score. It is based in the frequency of a keyword in the document, the inverse document frequency where rarer keywords get higher scores and the total weight of the keyword (eq. 7).

The Proposed Solution – Web Interface



Noticias

Una patadita de la buena suerte a Millonarios

Blog de usuario: La mesa de dinero Leía una columna de opinión de Juan Carlos Ortiz, un joven pero experimentado ex corredor de bolsa, de quién se dice es el hombre fuerte detrás de los grupos financieros Proyectar Valores e Interbolsa, retirado de la extinta Bolsa de Bogotá, y quien casi que de una manera inocente lanzó ayer la siguiente pregunta: "Y qué pasa si Millonarios va a la bolsa? leer más

Banco do Brasil aterrizará este año en Colombia

Está interesado en prestar sus servicios a las compañías brasileñas que están o harán presencia en el país. La llegada de Banco do Brasil a Colombia será una realidad este año, a través de la compra de una entidad del sistema financiero local o por medio de la apertura de una oficina de representación, que prestará sus servicios a las compañías brasileñas que están o harán presencia en el país. Así lo anunció el presidente de esta entidad, Carlos Massaru Takana-shi, quien llegó a Bogotá a anunciar una alianza con la Sociedad Administradora de Inversión del Grupo Interbolsa. leer más

InterBolsa expande su portafolio y será una fiduciaria

La administradora de inversiones del Grupo tramita la conversión a fiduciaria. Espera duplicar sus ingresos. Las inversiones en activos diferentes a lo que tradicionalmente ofrece el mercado de capitales será una de las apuestas del Grupo InterBolsa para los próximos meses. La construcción de un centro comercial en Santa Marta, la originación de hipotecas, la creación de un fondo de pensiones voluntarias y hasta las microfinanzas, son algunas de las iniciativas que ya están en trámite o que tienen entre los planes de mediano plazo. leer más

YAHOO! FINANCE


Enter Symbol SEARCH

Quote Web

NAME: EC NYQ LAST TRADE: \$58.71

Ecopetrol S.A. American Deposit

6 MONTHS CHART



NEWS

- Ecopetrol S.A. (EC): Zacks Rank Buy Zacks (Wed 1:00am)
- Top World Stocks Hit 50-Day Lines On Economic Worries

Personalization

- The Web is crawled for all of the defined assets
 - All interesting assets can be described and included
 - Interesting information sources can be defined as crawling seeds
 - Related interesting information sources are automatically discovered
- User defines his portfolio composition
 - Assets and amount of money in each personal investment alternative
 - The system calculates relative assets weights
- News and results are filtered and displayed considering personal investments
 - Filtering of personal information at query time

Tests and Results

- Fully implemented prototype
 - Debian 6
 - Java, Glassfish
 - Apache Hbase
 - Synesketch
- Data Extractor recollected data over 52 days
 - 38 working days for the stocks exchange data
 - 29919 financial news
 - 3152 elements from Social Media (mainly Twitter)
- Displayed information considers relationships through business domain for the actual user on query
- Information retrieval measures are taken over a sample of a hundred of documents, classified manually

Tests and Results

- Example for news content over 4 colombian enterprises:

<i>Company</i>	<i>Precision</i>	<i>Recall</i>
Odinsa	70%	84%
Interbolsa	28,5%	44%
Davivienda	75%	76%
Ecopetrol	68,9%	88%

<i>Company</i>	<i>Precision Positive Elements</i>	<i>Precision Negative Elements</i>	<i>Recall Positive Elements</i>	<i>Recall Negative Elements</i>
Odinsa	33%	39%	30%	40%
Interbolsa	67%	33%	20%	60%
Davivienda	41%	36%	30%	20%
Ecopetrol	39%	34%	50%	40%

- Comparable with the Hermes Framework and YourNews related works
- Good recall and precision for enterprises centered in one industry: Odinsa, Davivienda, Ecopetrol
- Poor precision for enterprises involved in several topics
- Synesketch sentiment analysis is not customizable, results are not domain-specific
 - LingPipe is a better option for this component, even if training is needed

Conclusions

- Successful integration of non-structured heterogeneous and domain-specific content
- More than 240 information sources
 - Multilanguage, international and regional coverage
 - Social media, specialized sources, news integration
 - Web and structured public data
- Specific domain model based on vector-space model
- Customized and configurable both in domain and user dimensions

Conclusions...

-
- Classification and information retrieval techniques in order to deliver appropriate content
 - Sentiment analysis and perception of financial information is included
 - NoSQL technology for scalability and flexibility
 - The proposed architecture can be used in other domain contexts
 - The semantic table can be defined for other domains



THANK YOU !

Questions?

Contact:

Camilo Restrepo-Arango
Arturo Henao-Chaparro
Claudia Jiménez-Guarín

Email: c.restrepo235@uniandes.edu.co
a.henao59@uniandes.edu.co
cjimenez@uniandes.edu.co

<http://comit.uniandes.edu.co>

Blog: <http://wordpress1.virtual.uniandes.edu.co>

