

Chapter 7

Use Case: Financial Information Integration

Financial information portals like Wall Street Journal Online¹, Bloomberg², Yahoo Finance³ and Google Finance⁴ enable investors to access a multitude of financial news, analyst reports, and postings from investment related discussion forums.

This chapter demonstrates how the Named Graphs data model and the Semantic Web Publishing Vocabulary could be used by a financial information portal to represent financial news, analyst reports, and newsgroup postings together with quality-related meta-information, such as provenance information, ratings, and background information about the information providers. The developed domain model will be used as a running example for explaining the capabilities of the WIQA - Information Quality Assessment Framework in the next part of this thesis.

7.1 Domain Model

This section describes a domain model for representing financial information together with quality-related meta-information as a set of named graphs. The domain model reuses existing terms from the Dublin Core [ISO03a], FOAF [BM04], RDF Schema [BG04] and ISO 3166-1 [ISO97] vocabularies. Domain specific terms are defined in the <http://www.fu-berlin/suhl/bizer/2006/FinVoc/> namespace, which is abbrevi-

¹<http://online.wsj.com/> (retrieved 09/25/2006)

²<http://quote.bloomberg.com/> (retrieved 09/25/2006)

³<http://finance.yahoo.com/> (retrieved 09/25/2006)

⁴<http://finance.google.com/> (retrieved 09/25/2006)

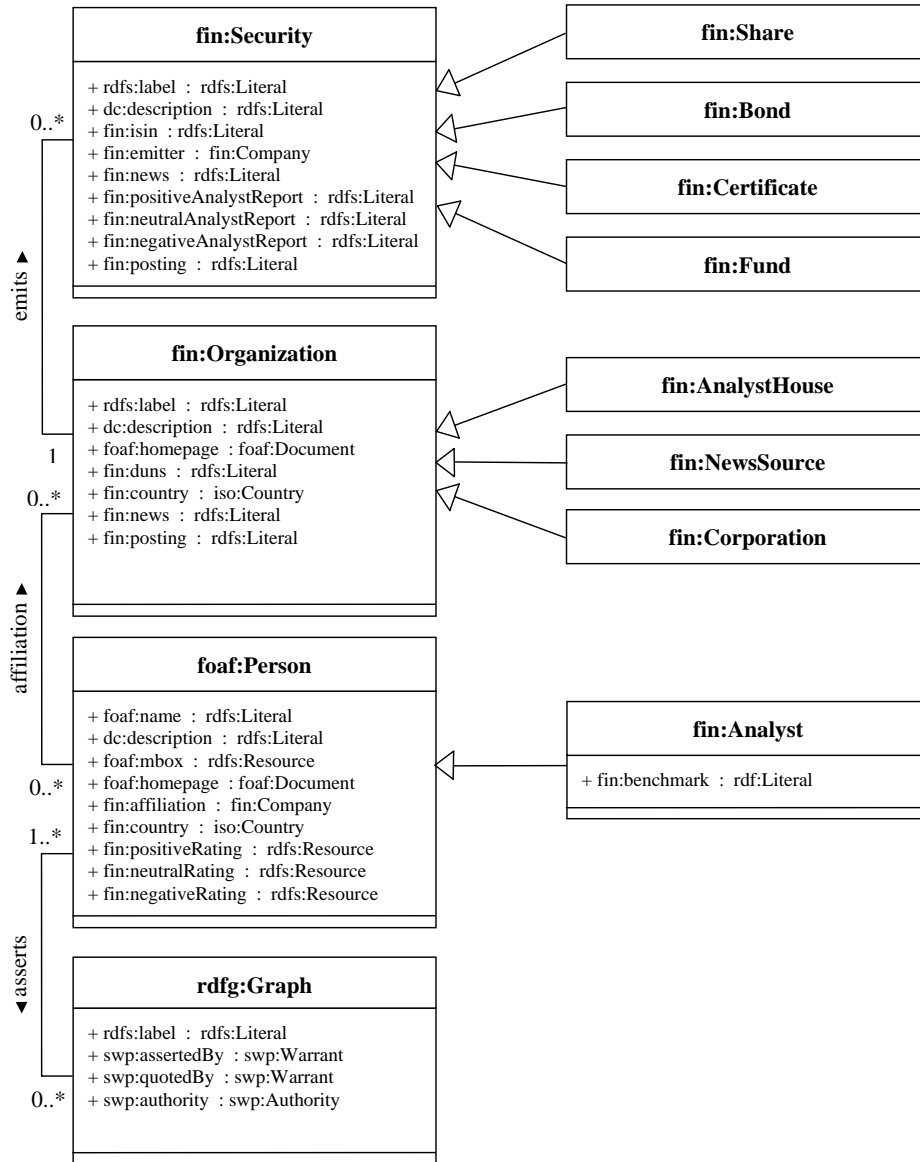


Figure 7.1: UML class diagram of the domain model.

ated using the `fin:` prefix. Figure 7.1 gives an overview about the classes of the model.

Securities are represented as instances of the class `fin:Security` and its subclasses `fin:Share`, `fin:Bond`, `fin:Fund`, `fin:Certificate`. Securities are identified by their International Securities Identifying Number (ISIN) [ISO01] which is mapped into the Universal Resource Name (URN) namespace [Moa97]. The name and a description of a security is represented using `rdfs:label` and `dc:description` properties. News, analyst reports, and discussion forum postings about a security are represented using the `fin:news`, `fin:positiveAnalystReport`, `fin:neutralAnalystReport`, `fin:negativeAnalystReport`, and `fin:posting` properties. Securities are related to the emitting organization by the `fin:emitter` property.

Organizations are represented as instances of the class `fin:Organization` and its subclasses `fin:Corporation`, `fin:AnalystHouse`, and `fin:NewsSource`. Organizations are identified using the Data Universal Numbering System (D-U-N-S) [Dun06] which is also mapped into the URN namespace. The name and the profile of an organization is represented using `rdfs:label` and `dc:description` properties. The `foaf:homepage` property contains a link to the main website of the organization. A ISO 3166-1 [ISO97] country code may be assigned to an organization using the `fin:country` property. News and discussion forum postings about organizations are represented using the `fin:news` and `fin:posting` properties.

Persons are represented as instances of the class `foaf:Person`. Persons are identified by their email address. The affiliation of a person to an `fin:Organization` is expressed using the `fin:affiliation` property. Persons may rate other persons, analysts, analyst houses, and news sources. A rating represents the subjective, overall impression of the rater about the quality of information that is published by the rated information provider. Ratings are expressed using the `fin:positiveRating`, `fin:neutralRating`, and `fin:negativeRating` properties. The class `fin:Analyst` is a subclass of `foaf:Person`. It consists of all persons working as analysts for a `fin:AnalystHouse`. `fin:Analysts` may have a `fin:benchmark` property which measures the accuracy of the recommendations of an analyst relative to his peers. The score is calculated using the StarMine algorithm [Yah06].

Named Graphs are typed as instances of the class `rdfg:Graph` [CBHS05a].

Provenance information about a graph is expressed using the Semantic Web Publishing Vocabulary described in Section 6.

7.2 Example Data Set

The domain model has been instantiated with example data from several publicly accessible financial information portals. The complete data set is available at <http://www.wiwiss.fu-berlin.de/suhl/bizer/wiqa/finUseCase/finData.trig>. The data was complemented with random ratings for analysts, analyst houses, and news sources which will be used in the next part of this thesis for demonstrating rating-based filtering policies.

Figures 7.2 and 7.3 show the TriG serialization of a subset of the collected data. The subset consists of seven named graphs. Each graph contains information from a specific information provider. The graphs `fd:GraphFromIntel` and `fd:GraphFromSiemens` contain general information about Intel and Siemens. The graph `fd:GraphFromPeterSmith` contains analyst reports about the stocks of both companies. The reports have been authored by Peter Smith, an analyst working for Deutsche Bank. The graph `fd:GraphFromMarkScott` contains two discussion forum posting authored by Mark Scott. The graph also contains information quality ratings for different sources. Mark Scott rates Peter Smith and Deutsche Bank positive and rates Dave Reynolds negative. The graph `fd:BackgroundInformation` contains background information about Peter Smith, Dave Reynolds, and Mark Scott. The first 5 triples of the graph assert that Peter Smith is a German analyst who works for Deutsche Bank and has a StarMine performance benchmark of 88. The next 4 triples contain background information about Deutsche Bank. `fd:GraphFromAggregator` contains provenance information about the other graphs. The first 3 triples, for instance, represent the information that `fd:GraphFromIntel` was asserted by Intel on 21st October 2005. Chapter 9 will describe how background and provenance information from the last two graphs can be used by different filtering policies.

```

1. @prefix rdf: <http://www.w3.org/1999/02/22-rdf-syntax-ns#> .
2. @prefix rdfs: <http://www.w3.org/2000/01/rdf-schema#> .
3. @prefix xsd: <http://www.w3.org/2001/XMLSchema#> .
4. @prefix dc: <http://purl.org/dc/elements/1.1/> .
5. @prefix foaf: <http://xmlns.com/foaf/0.1/> .
6. @prefix swp: <http://www.w3.org/2004/03/trix/swp-2/> .
7. @prefix iso: <http://www.daml.org/2001/09/countries/iso-3166-ont#> .
8. @prefix fin: <http://www.fu-berlin/suhl/bizer/2006/FinVoc/> .
9. @prefix fd: <http://www.fu-berlin/suhl/bizer/exampleDataset> .
10.
11. fd:GraphFromIntel {
12.
13.   <urn:x-DUNS:047897855> rdf:type fin:Corporation ;
14.     rdfs:label "Intel, Inc"^^xsd:string ;
15.     fin:country iso:US ;
16.     foaf:homepage <http://www.intel.com> .
17.   <urn:x-ISIN:US4581401001> rdf:type fin:Share ;
18.     fin:emitter <urn:x-DUNS:047897855> . }
19.
20. fd:GraphfromSiemens {
21.
22.   <urn:x-DUNS:316067164> rdf:type fin:Corporation ;
23.     rdfs:label "Siemens AG"^^xsd:string ;
24.     fin:country iso:DE ;
25.     foaf:homepage <http://www.siemens.com> .
26.   <urn:x-ISIN:DE0007236101> rdf:type fin:Share ;
27.     fin:emitter <urn:x-DUNS:316067164> . }
28.
29. fd:GraphFromPeterSmith {
30.
31.   <urn:x-ISIN:DE0007236101> fin:positiveAnalystReport "As Siemens
32.     agrees partnership with Novell unit SUSE ..."@EN .
33.   <urn:x-ISIN:US4581401001> fin:negativeAnalystReport "Chiphersteller
34.     Intel will nach Firmenangaben mit milliardenschweren ..."@DE . }
35.
36. fd:GraphFromJohnReynolds {
37.
38.   <urn:x-DUNS:316067164> fin:news "Siemens AG currently has no plans
39.     to sell its troubled mobile phone division..."@EN .
40.   <urn:x-DUNS:047897855> fin:news "Intel has record quarterly
41.     sales. Intel, the world's largest chipmaker..."@EN . }
42.
43. fd:GraphFromMarkScott {
44.
45.   <urn:x-ISIN:DE0007236101> fin:posting "I think the time ..."@EN .
46.   <urn:x-ISIN:US4581401001> fin:posting "As we have seen in ..."@EN .
47.   <mailto:mark@scott.com> fin:positiveRating <urn:x-DUNS:332907323> .
48.   <mailto:mark@scott.com> fin:positiveRating
49.     <mailto:peterSmith@deutsche-bank.de> .
50.   <mailto:mark@scott.com> fin:negativeRating
51.     <mailto:reynolds@ft.com> . }

```

Figure 7.2: Example graph set - Part 1.

```

52. fd:BackgroundInformation {
53.
54.   <mailto:peterSmith@deutsche-bank.de> rdf:type fin:Analyst ;
55.     foaf:name "Peter Smith"^^xsd:string ;
56.     fin:country iso:DE ;
57.     fin:affiliation <urn:x-DUNS:332907323> ;
58.     fin:benchmark "83"^^xsd:integer .
59. <urn:x-DUNS:332907323> rdf:type fin:AnalystHouse ;
60.   foaf:name "Deutsche Bank"^^xsd:string ;
61.   foaf:homepage <http://www.db.com> ;
62.   fin:country iso:DE .
63. <mailto:reynolds@ft.com> rdf:type fin:Analyst ;
64.   foaf:name "John Reynolds"^^xsd:string ;
65.   fin:country iso:UK ;
66.   fin:affiliation <urn:x-DUNS:42307553> .
67. <urn:x-DUNS:42307553> rdf:type fin:NewsSource ;
68.   foaf:name "Financial Times"^^xsd:string ;
69.   foaf:homepage <http://news.ft.com/> ;
70.   fin:country iso:UK .
71. <mailto:mark@scott.com> rdf:type foaf:Person ;
72.   foaf:name "Mark Scott"^^xsd:string ;
73.   fin:country iso:UK . }
74.
75. fd:GraphFromAggregator {
76.
77.   fd:GraphFromIntel swp:assertedBy fd:IntelWarrant .
78.     fd:IntelWarrant swp:authority <urn:x-DUNS:047897855> ;
79.     dc:date "2005-10-21"^^xsd:date .
80.   fd:GraphFromSiemens swp:assertedBy fd:SiemensWarrant .
81.     fd:SiemensWarrant swp:authority <urn:x-DUNS:316067164> ;
82.     dc:date "2005-08-21"^^xsd:date .
83.   fd:GraphFromPeterSmith swp:assertedBy fd:PsWarrant .
84.     fd:PsWarrant swp:authority <mailto:peterSmith@deutsche-bank.de> ;
85.     dc:date "2005-11-20T12:40:44"^^xsd:dateTime .
86.   fd:GraphFromJohnReynolds swp:assertedBy fd:JrWarrant .
87.     fd:JrWarrant swp:authority <mailto:reynolds@ft.com> ;
88.     dc:date "2005-11-20T09:44:32"^^xsd:dateTime .
89.   fd:GraphFromMarkScott swp:assertedBy fd:MsWarrant ;
90.     fd:MsWarrant swp:authority <mailto:mark@scott.com> ;
91.     dc:date "2005-11-20T17:22:10"^^xsd:dateTime .
92.   fd:fd:BackgroundInformation swp:assertedBy fd:GraphFromAggregator .
93.   fd:GraphFromAggregator swp:assertedBy fd:GraphFromAggregator ;
94.     swp:authority <urn:x-DUNS:377667161> ;
95.     dc:date "2005-11-21T18:20:00"^^xsd:dateTime . }

```

Figure 7.3: Example graph set - Part 2.