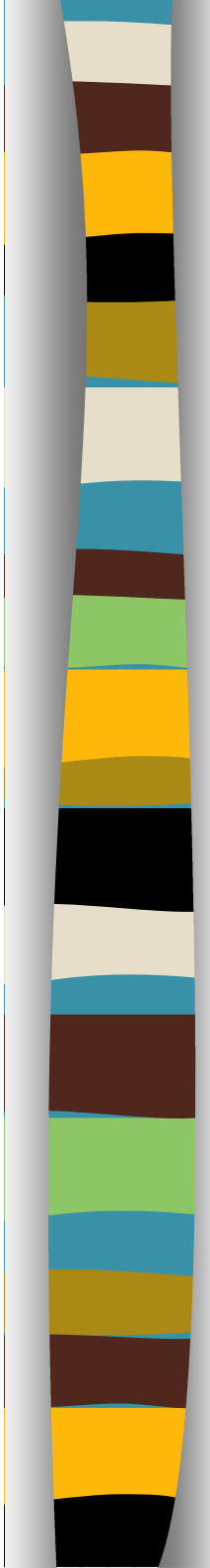
Meta-Stars: Multidimensional Modeling for Social Business Intelligence

Enrico Gallinucci - Matteo Golfarelli - Stefano Rizzi

University of Bologna - Italy



Summary

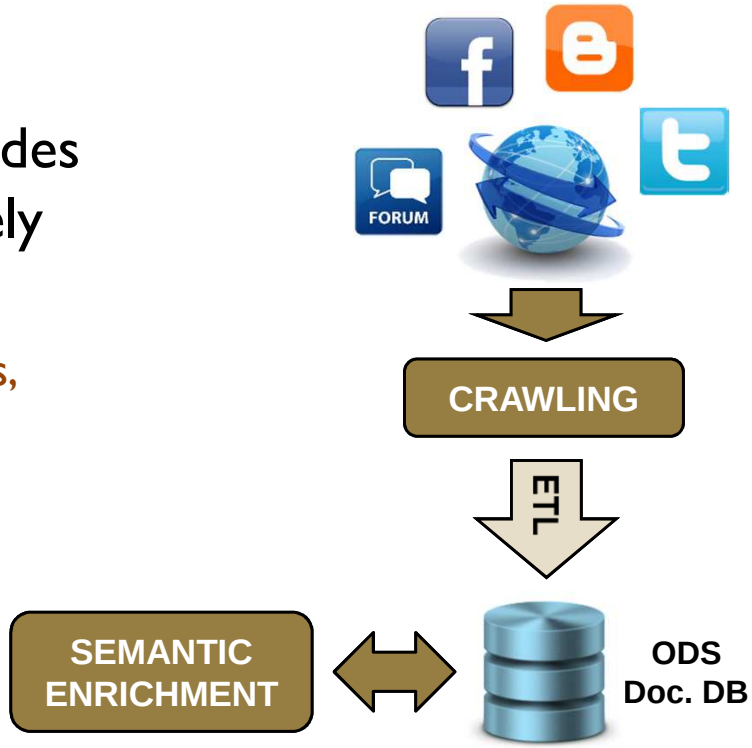
- Introduction: Social BI
- Topic hierarchy schema
- The Meta-Star approach
- Querying Meta-Stars
- Evaluation
- Conclusions

Social BI

- The enormous amount of UGC provides decision makers with a fresh and timely perception of the market's mood
  - People voluntarily share tastes and thoughts, influencing business and society

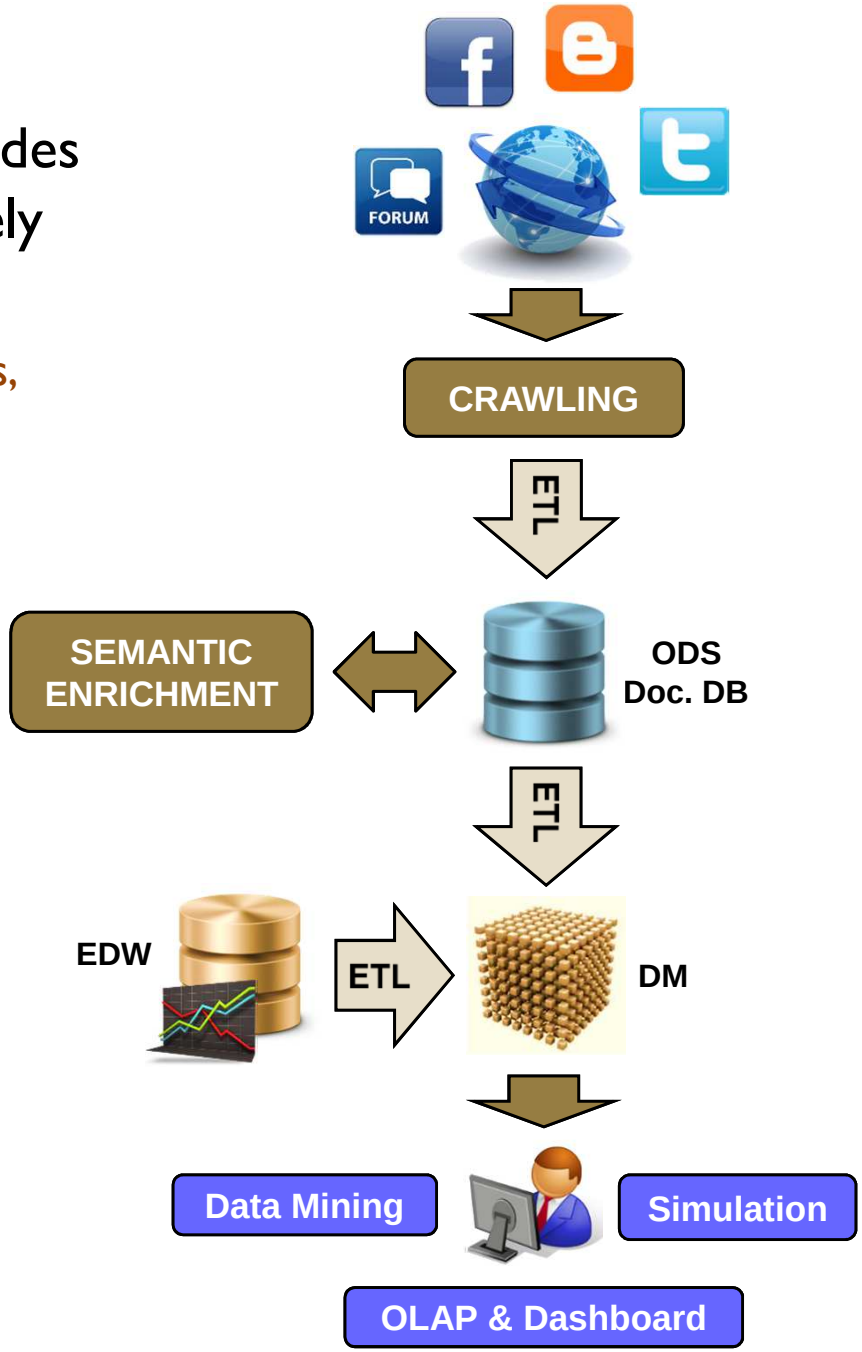
Social BI

- The enormous amount of UGC provides decision makers with a fresh and timely perception of the market's mood
  - People voluntarily share tastes and thoughts, influencing business and society
- Commercial tools are capable of digging useful information, identifying relevant parts and assigning them a polarity



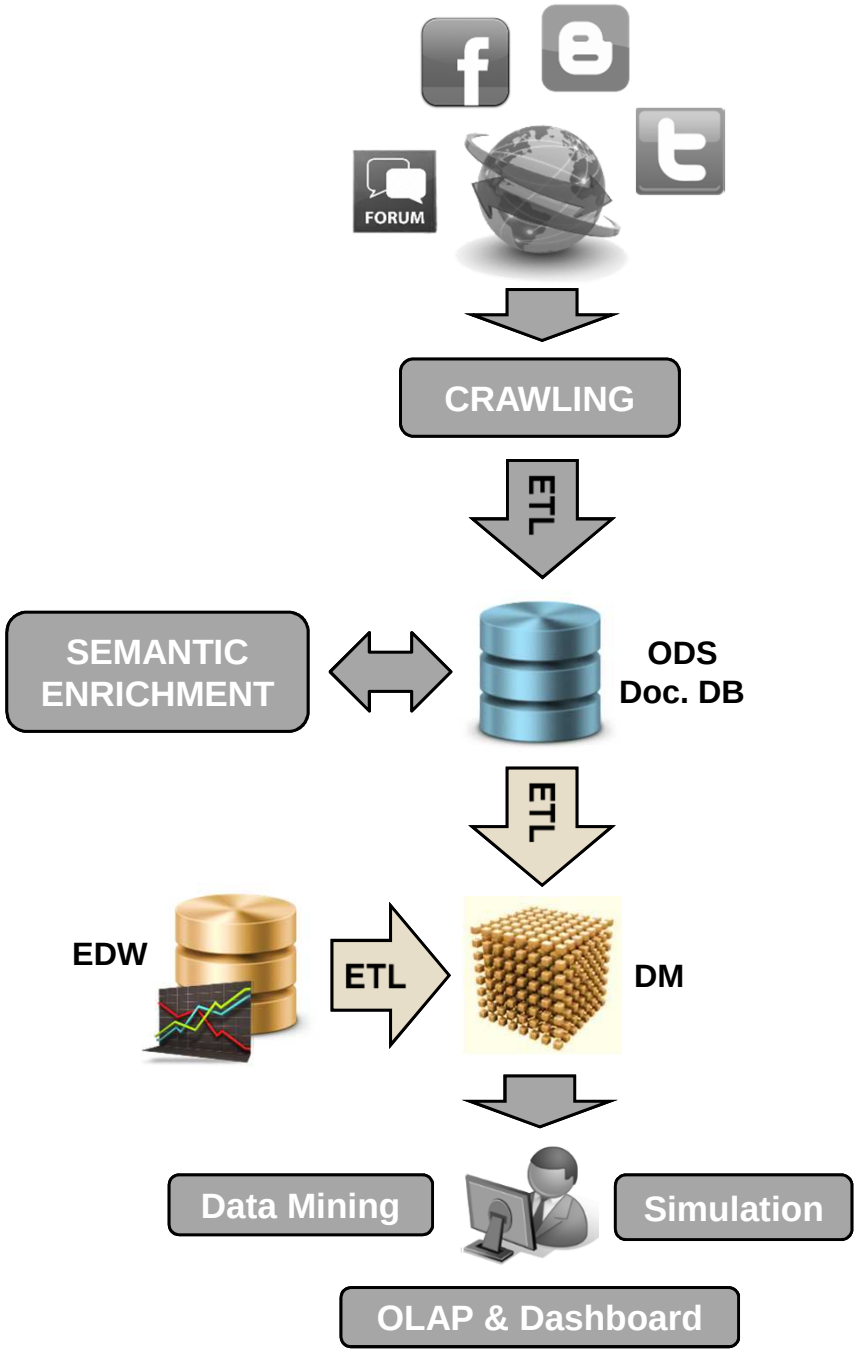
Social BI

- The enormous amount of UGC provides decision makers with a fresh and timely perception of the market's mood
  - People voluntarily share tastes and thoughts, influencing business and society
- Commercial tools are capable of digging useful information, identifying relevant parts and assigning them a polarity
- Current topic of research is the integration of these unstructured data to the Enterprise DW
  - Analysis capabilities of current commercial tools are very limited



Social BI

- Our work is focused on the individuation of a smart model to best represent social data
  - Allow powerful analysis
  - Easy integration with enterprise data
  - Handle heterogeneity and dynamicity of social data



Analysis of textual UGC

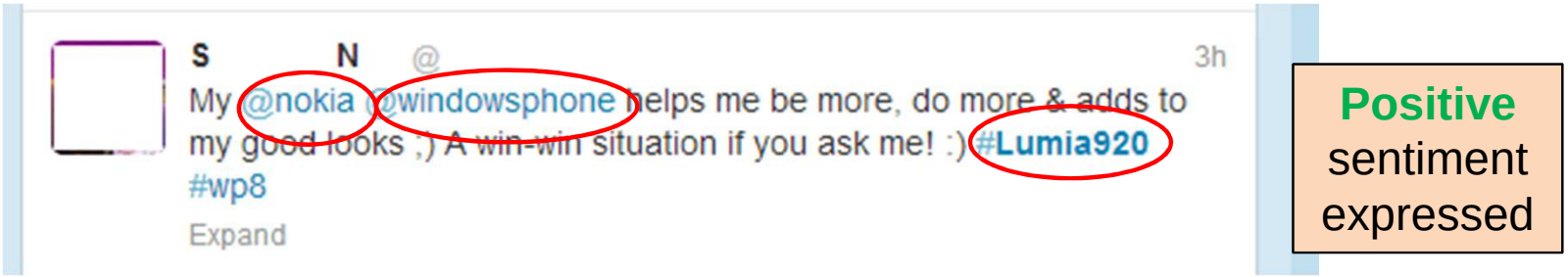
- A key role in the analysis is played by topics, meant as specific concepts of interest within the subject area

Analysis of textual UGC

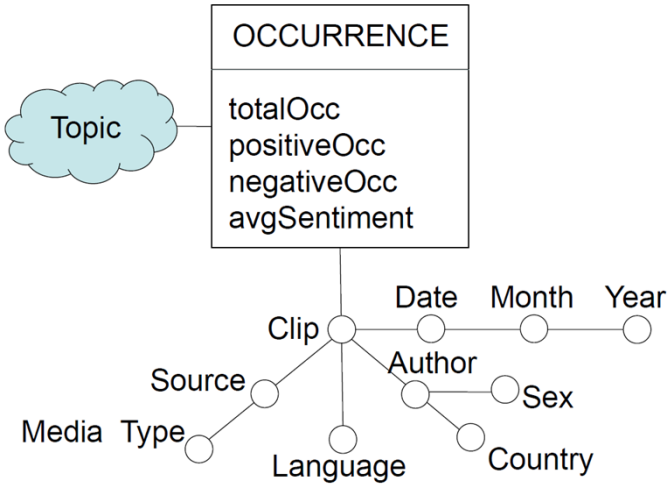
- A key role in the analysis is played by topics, meant as specific concepts of interest within the subject area

Analysis of textual UGC

- A key role in the analysis is played by topics, meant as specific concepts of interest within the subject area

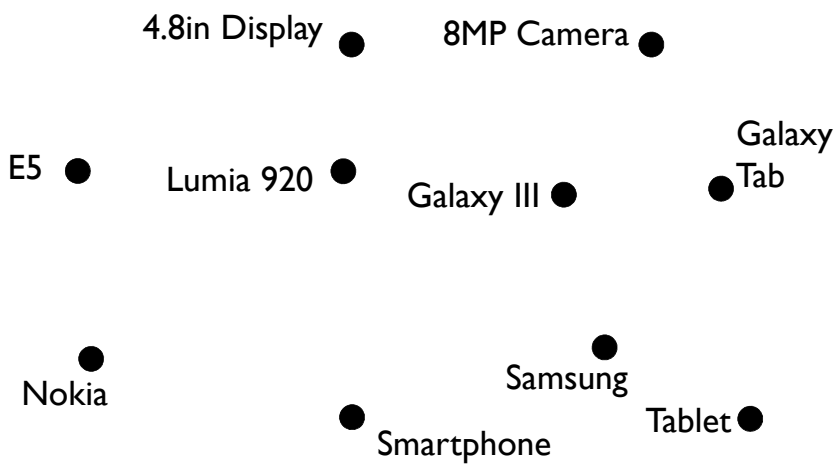


- Topics are an obvious candidate dimension of the cubes for Social BI, but:
  - Trending topic are heterogeneous and change quickly over time
  - A classical dimension table with a static hierarchy is not suitable



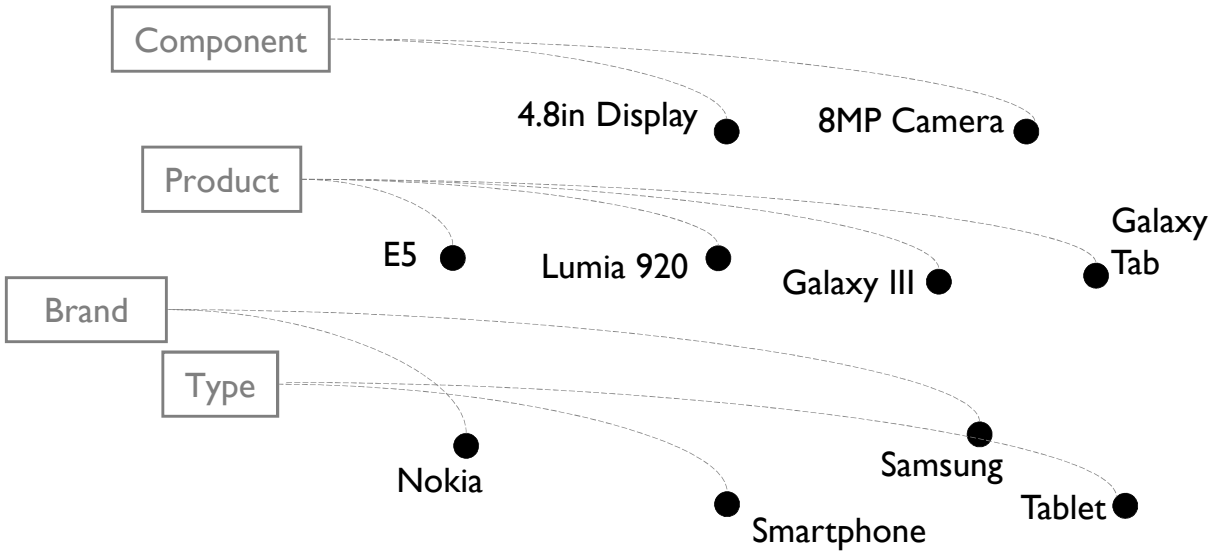
Topic hierarchy schema

- Consider a mobile-oriented scenario



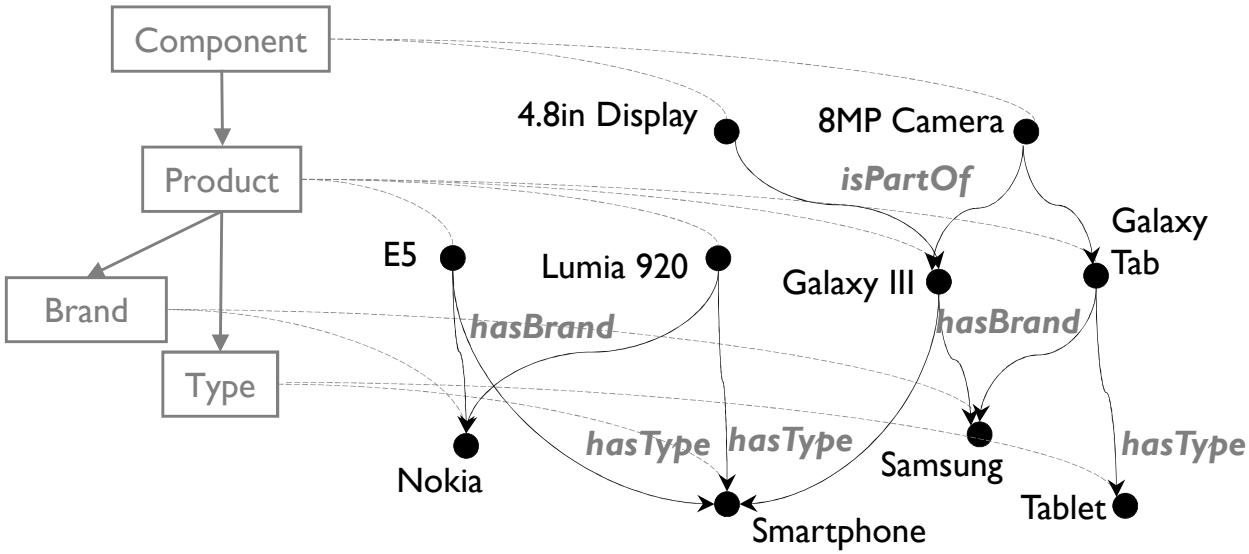
Topic hierarchy schema

- Consider a mobile-oriented scenario
  - Most topics can be classified into **levels**, that correspond to aggregation levels in traditional hierarchies



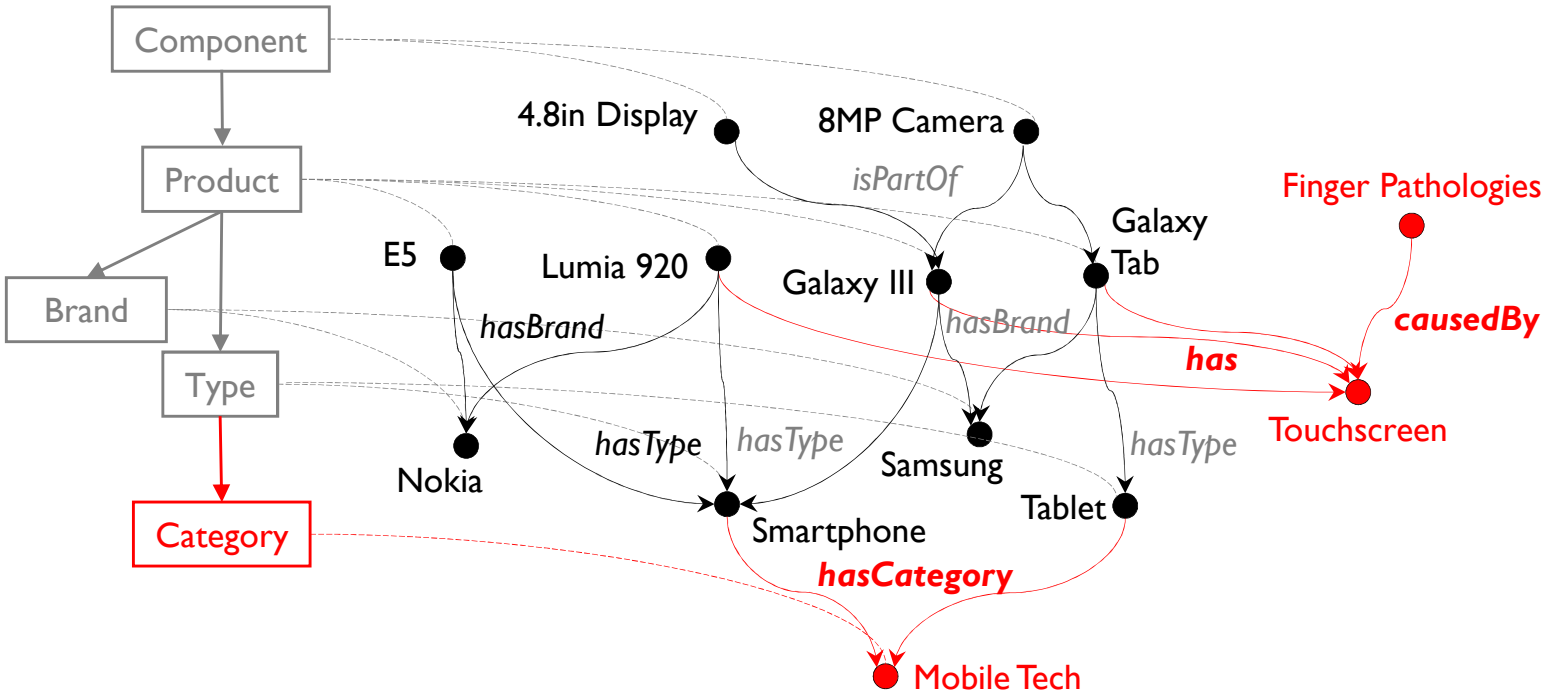
Topic hierarchy schema

- Consider a mobile-oriented scenario
  - Most topics can be classified into **levels**, that correspond to aggregation levels in traditional hierarchies
  - Relationships between topics highlight roll-up relationships



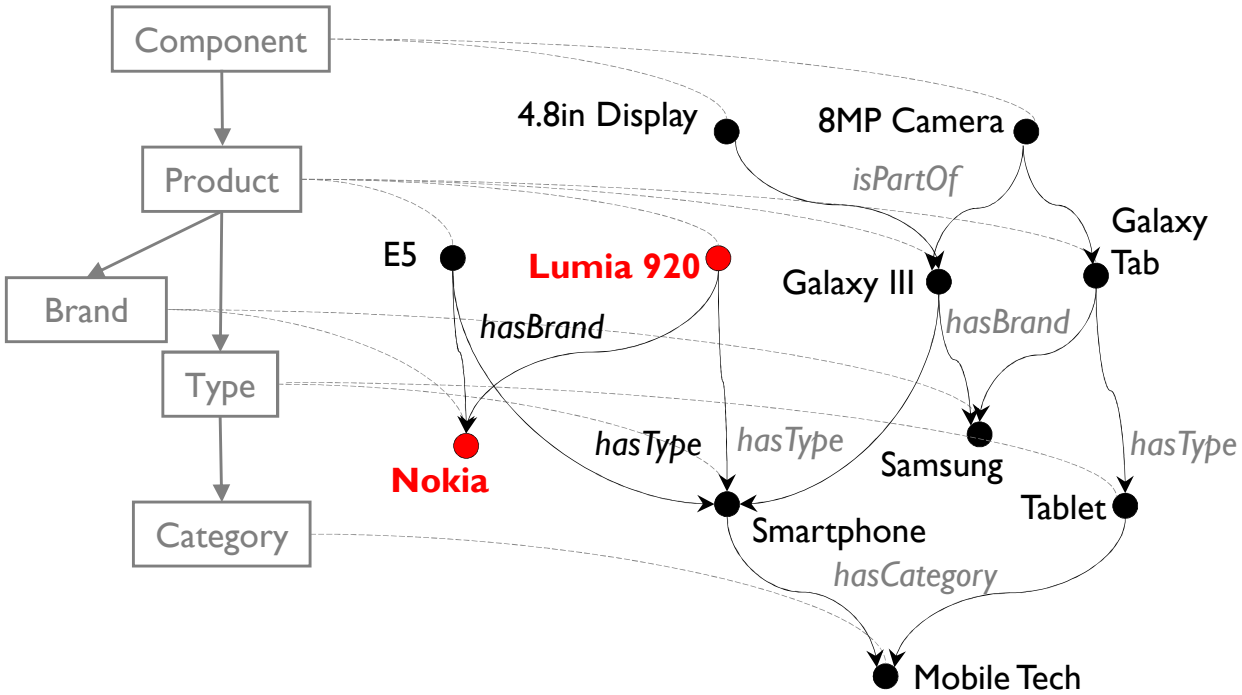
Topic hierarchy schema

- How is a topic hierarchy different from a traditional hierarchy?
  1. Dynamicity: new topics, relationships and aggregation levels might be added at any time



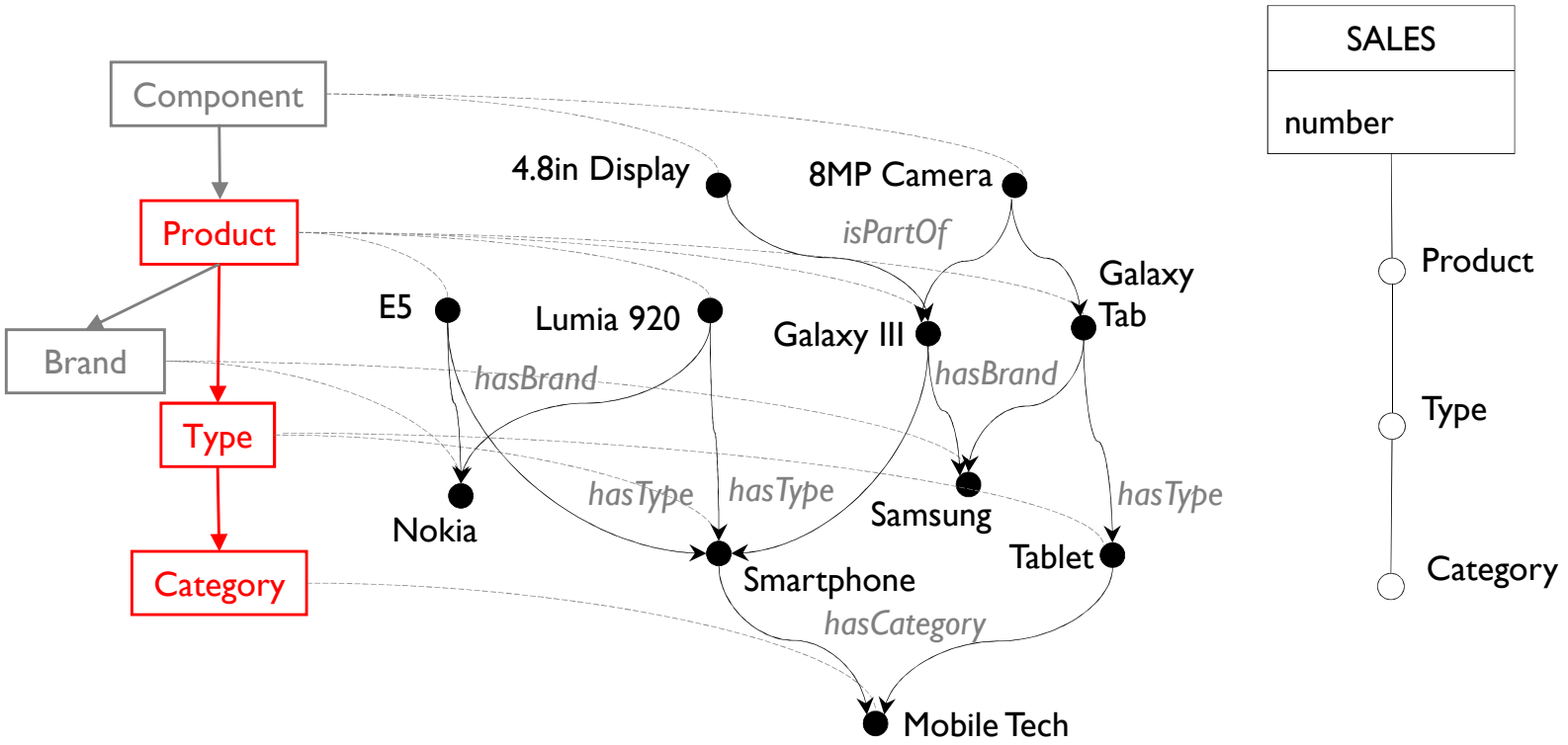
Topic hierarchy schema

- How is a topic hierarchy different from a traditional hierarchy?
  1. Dynamicity: new topics, relationships and aggregation levels might be added at any time
  2. Mixed granularity (facts associated to non leaf-topics) and unbalanced hierarchies



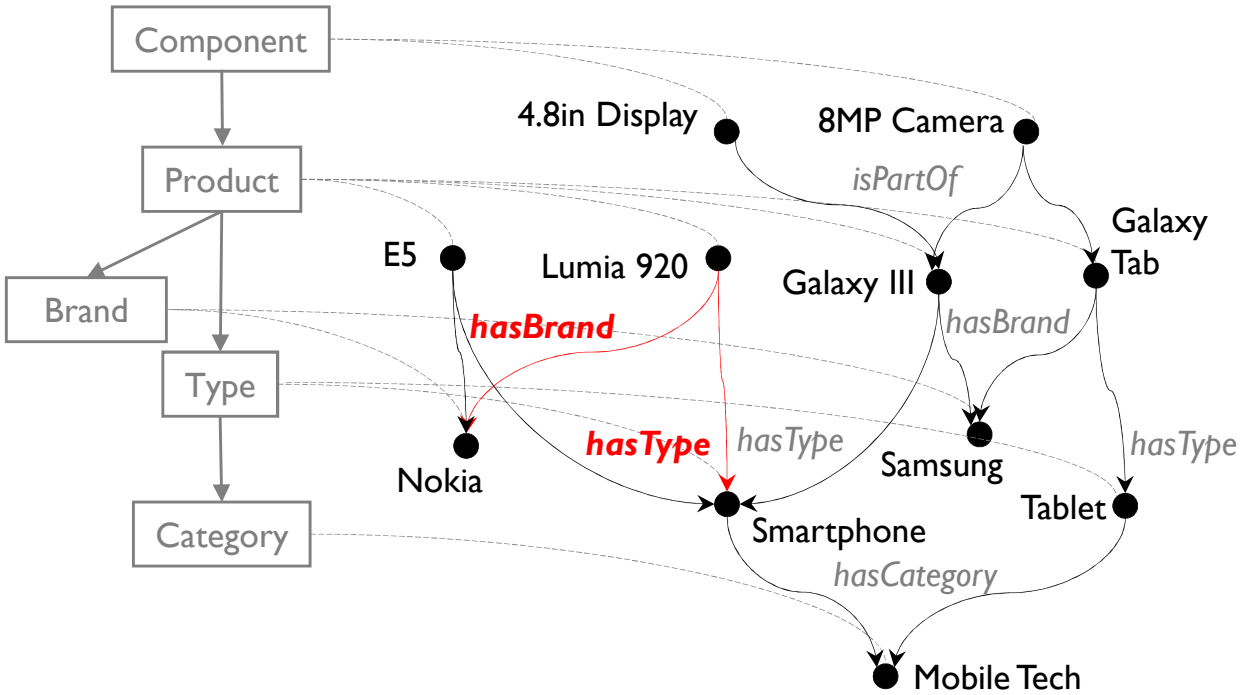
Topic hierarchy schema

- How is a topic hierarchy different from a traditional hierarchy?
  1. Dynamicity: new topics, relationships and aggregation levels might be added at any time
  2. Mixed granularity (facts associated to non leaf-topics) and unbalanced hierarchies
  3. Integration: some topics are also part of business hierarchies and require a direct connection with the enterprise cube



Topic hierarchy schema

- How is a topic hierarchy different from a traditional hierarchy?
  1. Dynamicity: new topics, relationships and aggregation levels might be added at any time
  2. Mixed granularity (facts associated to non leaf-topics) and unbalanced hierarchies
  3. Integration: some topics are also part of business hierarchies and require a direct connection with the enterprise cube
  4. Semantics: roll-up relationships between topics can have different semantics



The Meta-Star approach

- Meta-Stars overcome these issues by using a combination of modeling strategies

- Navigation tables
  - Support hierarchy instances with different lengths and non-leaf facts
  - Allow different roll-up semantics to be explicitly annotated

- Meta-modeling
  - Enable hierarchy heterogeneity and dynamicity to be accommodated

- Traditional dimension tables
  - Easy integration with standard business hierarchies

The Meta-Star approach

- Implementation of a Meta-Star requires two components:

1. A Topic Table
   - Stores all the topics of the hierarchy
   - Topic levels can be modeled in a *static* way (i.e., like in a classical dimension table)

2. A Rollup Table
   - Stores every relationship between two topics in the transitive closure

The Meta-Star approach

- Implementation of a Meta-Star: the **topic table**
  - One row for each topic

TOPIC_T

IdT	Topic	Level
1	8MP Camera	Component
2	Galaxy III	Product
3	Galaxy Tab	Product
4	Smartphone	Type
5	Tablet	Type
6	Mobile Tech	Category
7	Samsung	Brand
8	Finger Path.	-
9	Touchscreen	-
...	...	...

The Meta-Star approach

- Implementation of a Meta-Star: the **topic table**
  - One row for each topic

TOPIC_T

IdT	Topic	Level
1	8MP Camera	Component
2	Galaxy III	**Product**
3	Galaxy Tab	**Product**
4	Smartphone	**Type**
5	Tablet	**Type**
6	Mobile Tech	**Category**
7	Samsung	Brand
8	Finger Path.	-
9	Touchscreen	-
...	...	...

The Meta-Star approach

- Implementation of a Meta-Star: the **topic table**
  - One row for each topic
  - Columns for each static level, like in a classical dimension table

TOPIC_T

IdT	Topic	Level	**Product**	**Type**	**Category**
1	8MP Camera	Component	-	-	-
2	Galaxy III	Product	Galaxy III	Smartphone	Mobile Tech
3	Galaxy Tab	Product	Galaxy Tab	Tablet	Mobile Tech
4	Smartphone	Type	-	Smartphone	Mobile Tech
5	Tablet	Type	-	Tablet	Mobile Tech
6	Mobile Tech	Category	-	-	Mobile Tech
7	Samsung	Brand	-	-	-
8	Finger Path.	-	-	-	-
9	Touchscreen	-	-	-	-
…	…	…	…	…	…

The Meta-Star approach

- Implementation of a Meta-Star: the **roll-up table**
  - One row for each topic
  - One row for each arc in the transitive closure of the hierarchy

ROLLUP_T

ChildId	RollUpSignature	FatherId
1	000000	1
2	000000	2
…	000000	…
1	100000	2
1	100000	3
2	010000	4
2	001000	7
4	000100	6
8	000001	9
2	000010	9
…	…	…
1	110000	4
1	110000	5
1	101000	7
1	100010	9
2	010100	6
3	010100	6
…	…	…
1	110100	6
…	…	…

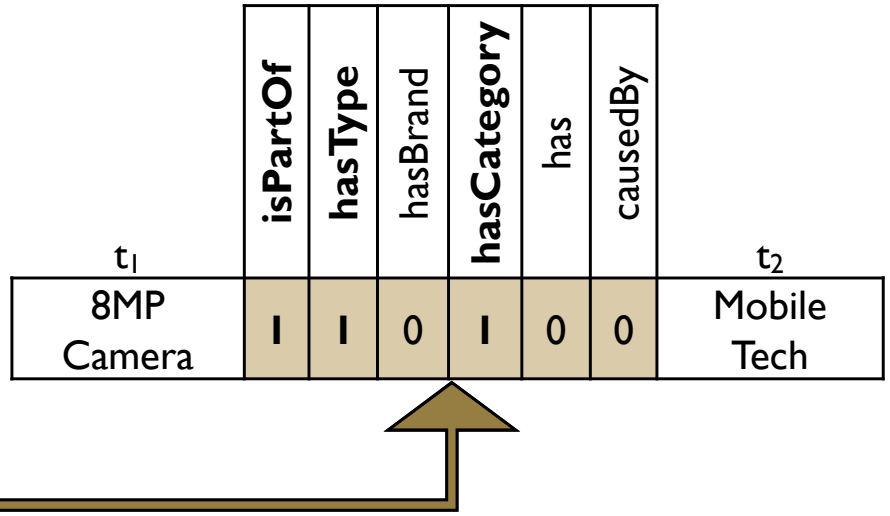
The Meta-Star approach

- Implementation of a Meta-Star: the **roll-up table**
  - One row for each topic
  - One row for each arc in the transitive closure of the hierarchy

ROLLUP_T

ChildId	RollUpSignature	FatherId
1	000000	1
2	000000	2
…	000000	…
1	100000	2
1	100000	3
2	010000	4
2	001000	7
4	000100	6
8	000001	9
2	000010	9
…	…	…
1	110000	4
1	110000	5
1	101000	7
1	100010	9
2	010100	6
3	010100	6
…	…	…
1	110100	6
…	…	…

- Each bit of the *roll-up signature* corresponds to one roll-up semantics
- If the hierarchy includes a directed path from t_1 to t_2, the bits corresponding to the involved roll-up semantics are set to 1



The Meta-Star approach

- The combination of meta-modeling with the roll-up table accommodates the dynamicity of the topic hierarchy

TOPIC_T

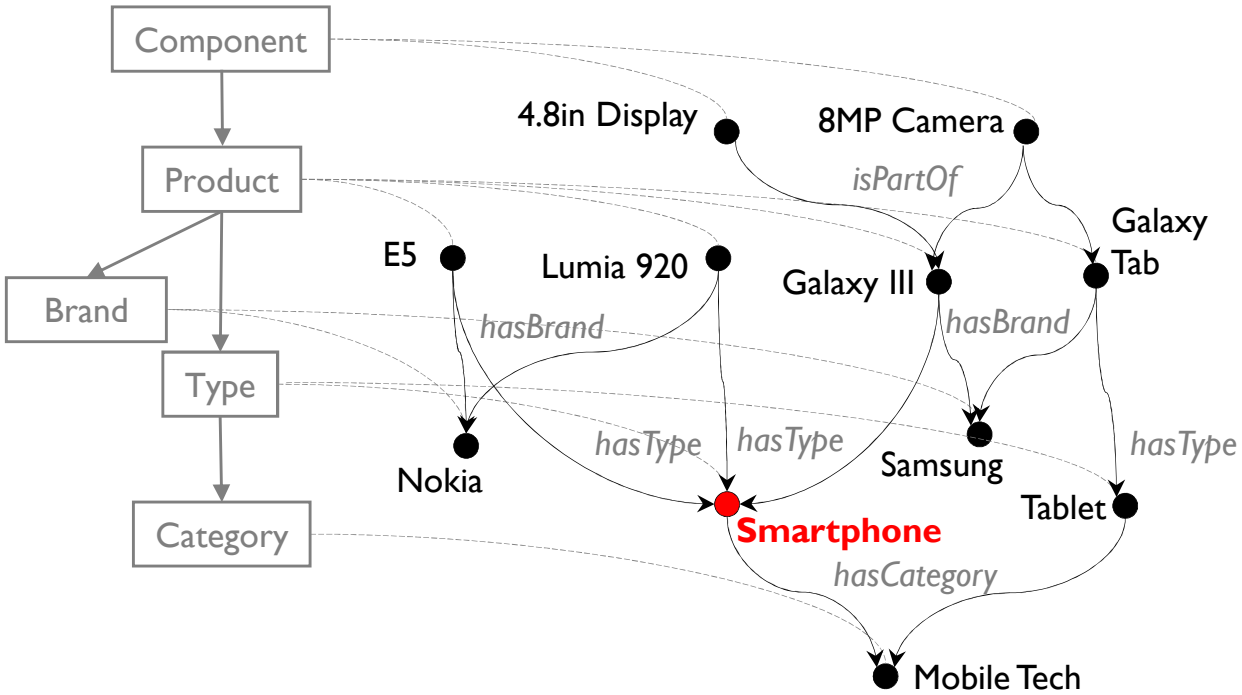
IdT	Topic	Level
1	8MP Camera	Component
2	Galaxy III	Product
3	Galaxy Tab	Product
4	Smartphone	Type
5	Tablet	Type
6	**Mobile Tech**	**Category**
7	Samsung	Brand
8	Finger Path.	-
9	Touchscreen	-
...	...	...

ROLLUP_T

ChildId	RollUpSignature	FatherId
1	000 **0** 00	1
2	000 **0** 00	2
…	000 **0** 00	…
1	100 **0** 00	2
1	100 **0** 00	3
2	010 **0** 00	4
2	001 **0** 00	7
4	**000 1 00**	**6**
8	000 **0** 01	9
2	000 **0** 10	9
…	…	…
1	110 **0** 00	4
1	110 **0** 00	5
1	101 **0** 00	7
1	100 **0** 10	9
2	**010 1 00**	**6**
3	**010 1 00**	**6**
…	…	…
1	**110 1 00**	**6**
…	…	…

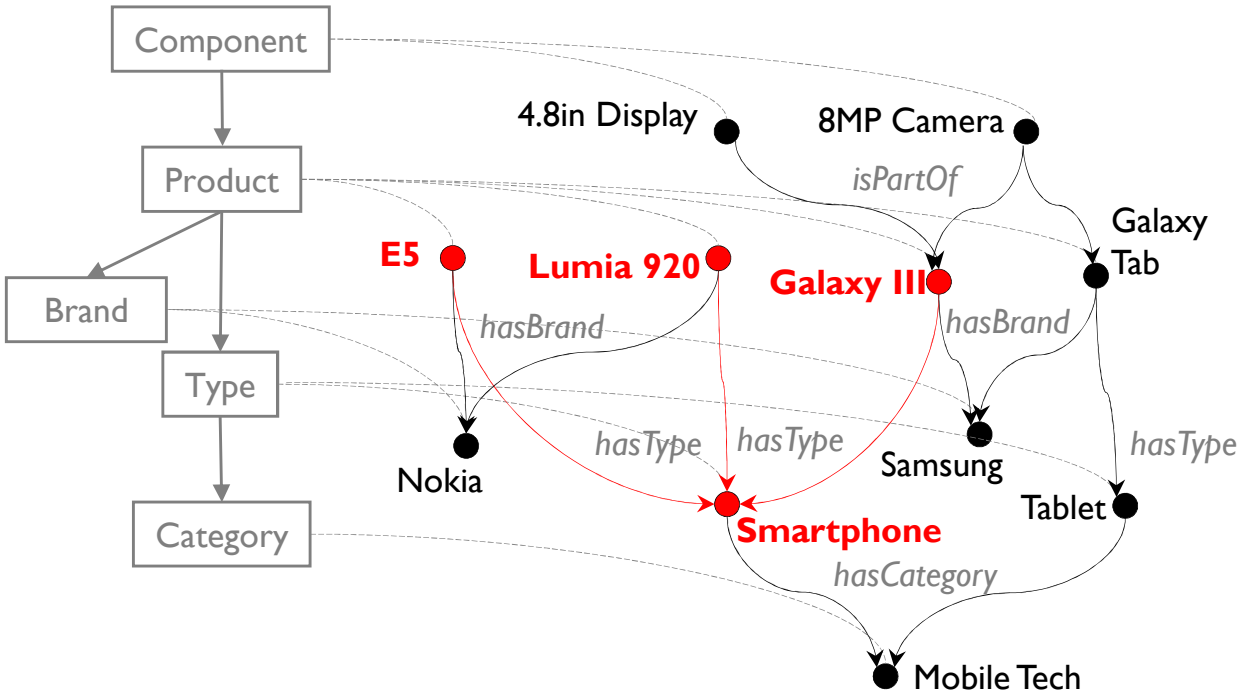
Querying Meta-Stars

- Question: what is the current *average sentiment* over smartphones?
  - Facts can be associated to non-leaf topics
  - Result's meaning is highly influenced by the involved semantics



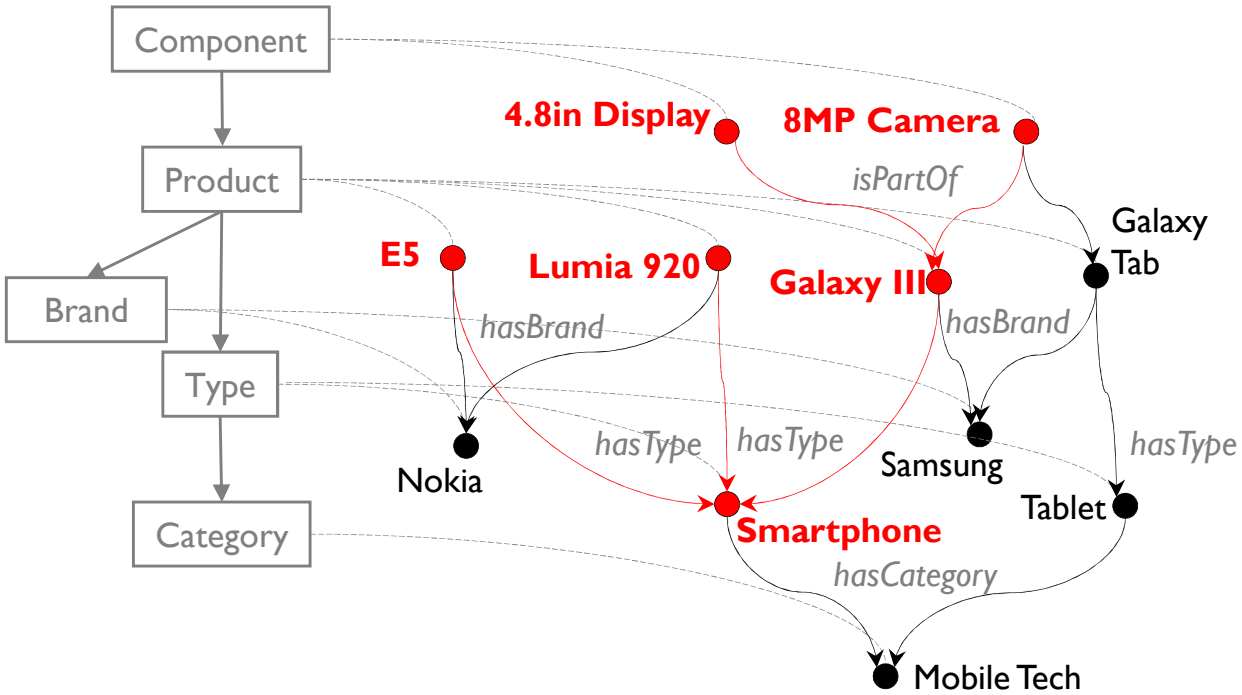
Querying Meta-Stars

- Question: what is the current *average sentiment* over smartphones?
  - Facts can be associated to non-leaf topics
  - Result's meaning is highly influenced by the involved semantics



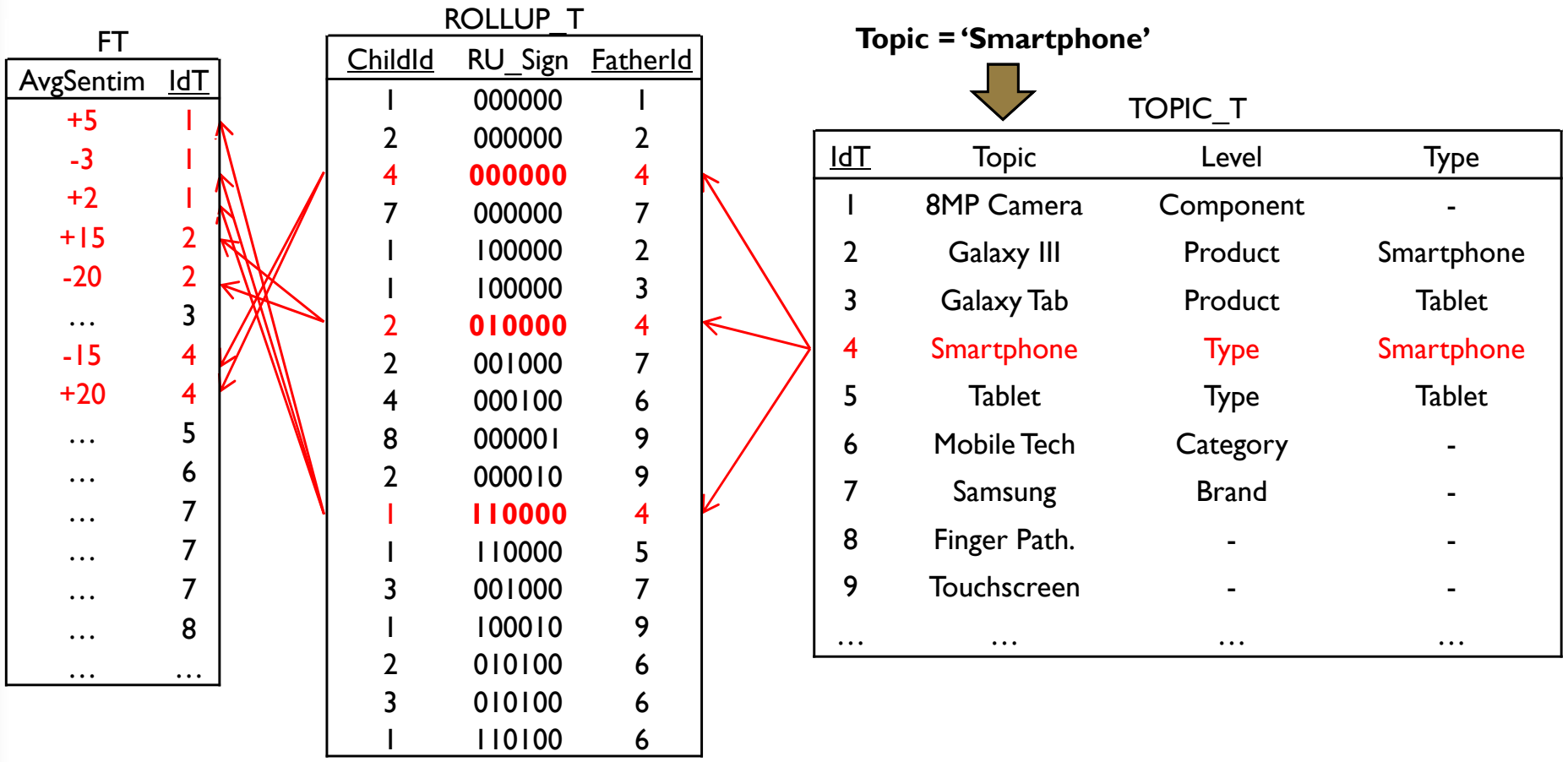
Querying Meta-Stars

- Question: what is the current *average sentiment* over smartphones?
  - Facts can be associated to non-leaf topics
  - Result's meaning is highly influenced by the involved semantics



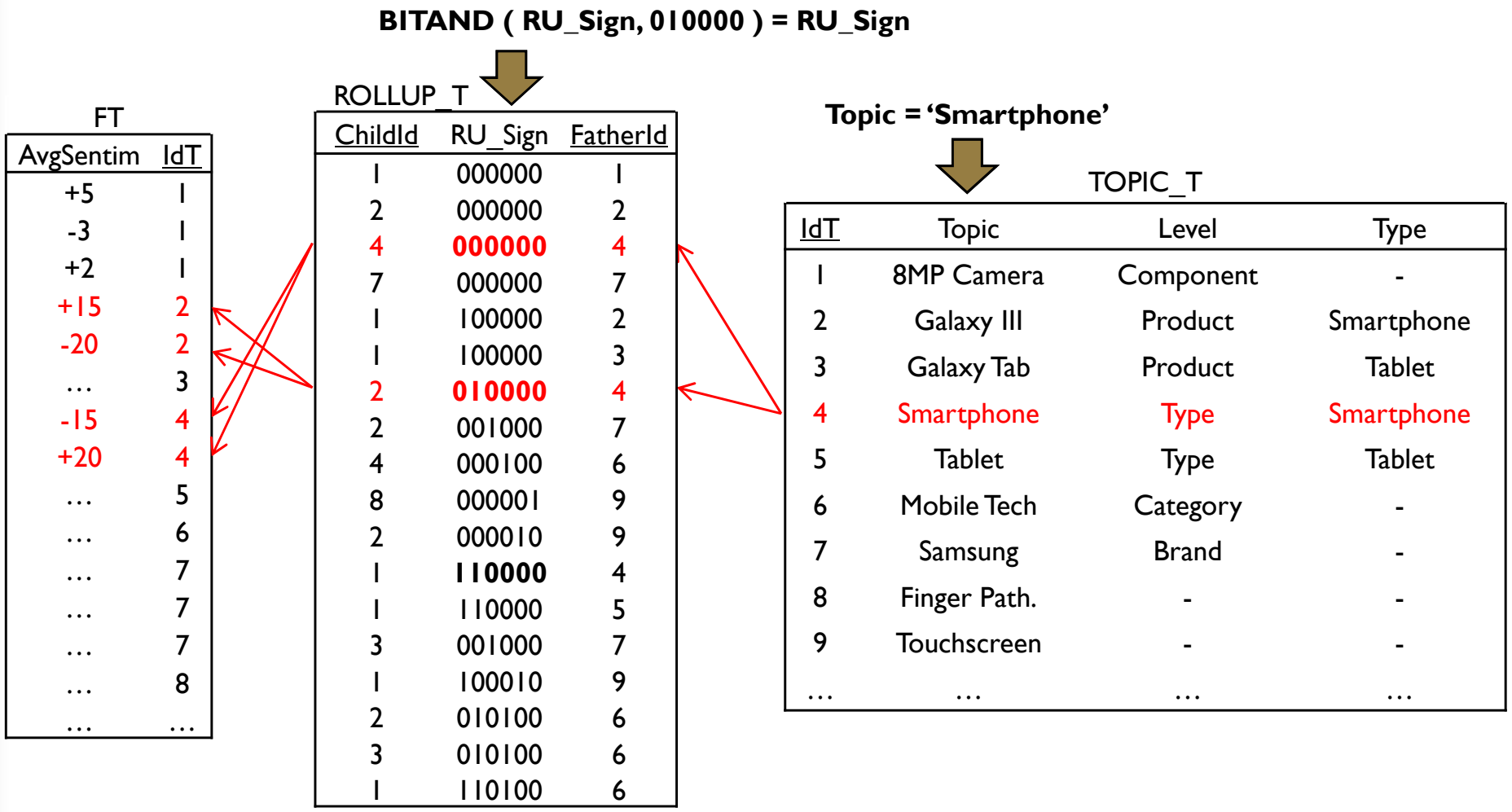
Querying Meta-Stars

- Question: what is the current *average sentiment* over smartphones?
  - Example of query with topic aggregation



Querying Meta-Stars

- Question: what is the current *average sentiment* over smartphones?
  - Example of query with **semantic-aware** topic aggregation



FT

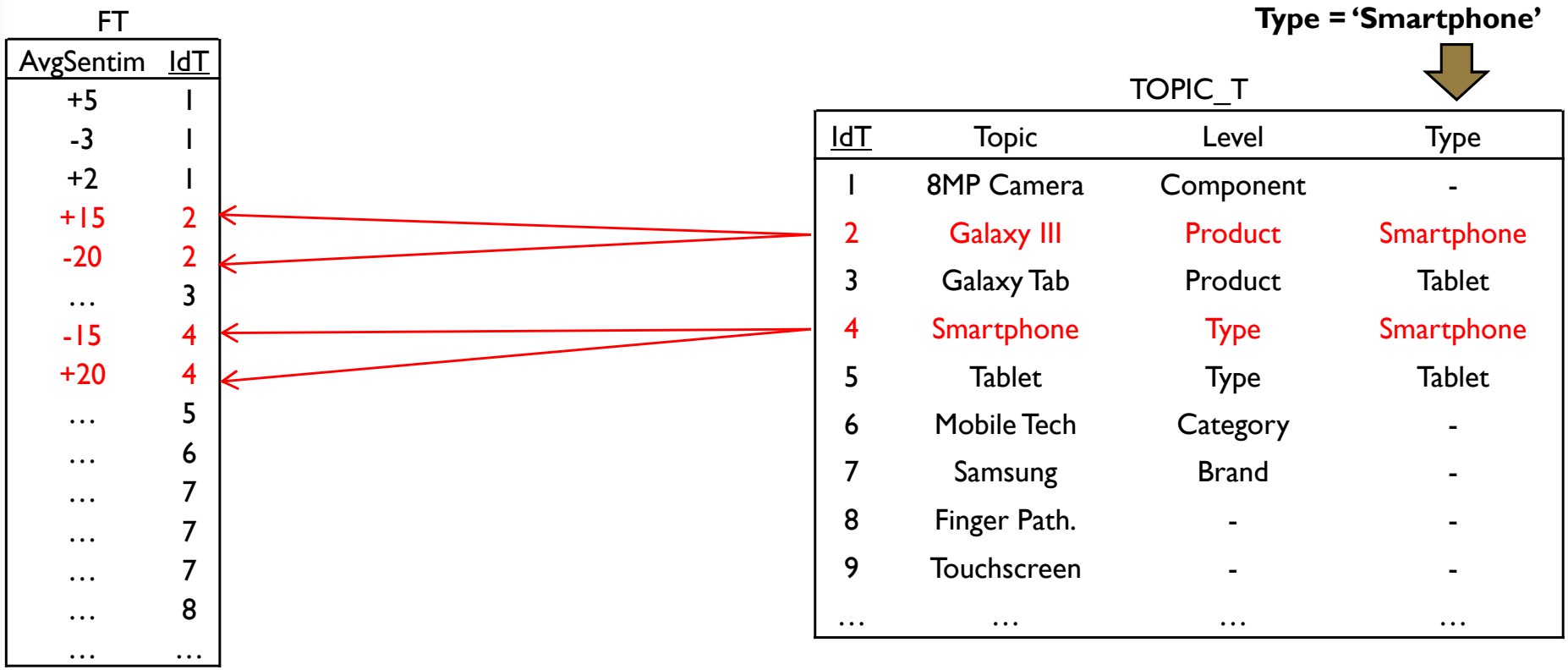
AvgSentim	IdT
+5	1
-3	1
+2	1
+15	2
-20	2
…	3
-15	4
+20	4
…	5
…	6
…	7
…	7
…	7
…	8
…	…

ChildId	RU_Sign	FatherId
1	000000	1
2	000000	2
4	**000000**	4
7	000000	7
1	100000	2
1	100000	3
2	**010000**	4
2	001000	7
4	000100	6
8	000001	9
2	000010	9
1	**110000**	4
1	110000	5
3	001000	7
1	100010	9
2	010100	6
3	010100	6
1	110100	6

IdT	Topic	Level	Type
1	8MP Camera	Component	-
2	Galaxy III	Product	Smartphone
3	Galaxy Tab	Product	Tablet
4	Smartphone	Type	Smartphone
5	Tablet	Type	Tablet
6	Mobile Tech	Category	-
7	Samsung	Brand	-
8	Finger Path.	-	-
9	Touchscreen	-	-
…	…	…	…

Querying Meta-Stars

- Question: what is the current *average sentiment* over smartphones?
  - Example of query with topic aggregation using static levels



Evaluation

- Performances of Meta-Star are compared with traditional star schemata using queries with Topic Aggregation

Topic hier.	|Group-by|	FT1		FT2	
		Meta-star	Star s.	Meta-star	Star s.
H1	0				
	1				
	2				
H2	0				
	1				
	2				
H3	0				
	1				
	2				

Evaluation

- Performances of Meta-Star are compared with traditional star schemata using queries with Topic Aggregation

FT1 → 1M facts
FT2 → 10M facts

Topic hier.	|Group-by|	FT1		FT2	
		Meta-star	Star s.	Meta-star	Star s.
H1	0				
	1				
	2				
H2	0				
	1				
	2				
H3	0				
	1				
	2				

Evaluation

- Performances of Meta-Star are compared with traditional star schemata using queries with Topic Aggregation

FT1 → 1M facts
FT2 → 10M facts

Topic hier.	|Group-by|	FT1		FT2	
		Meta-star	Star s.	Meta-star	Star s.
H1	0				
	1				
	2				
H2	0				
	1				
	2				
H3	0				
	1				
	2				

Topic hier.	| TOPIC_T |	| ROLLUP_T |	fan-out	tree-height
H1	106	626	4	4
H2	658	4514	8	4
H3	27,306	334,962	4	8

Evaluation

- Performances of Meta-Star are compared with traditional star schemata using queries with Topic Aggregation

Number of levels in the group-by predicate

FT1 → 1M facts
FT2 → 10M facts

Topic hier.	|Group-by|	FT1		FT2	
		Meta-star	Star s.	Meta-star	Star s.
H1	0				
	1				
	2				
H2	0				
	1				
	2				
H3	0				
	1				
	2				

Topic hier.	| TOPIC_T |	| ROLLUP_T |	fan-out	tree-height
H1	106	626	4	4
H2	658	4514	8	4
H3	27,306	334,962	4	8

Evaluation

- Performances of Meta-Star are compared with traditional star schemata using queries with Topic Aggregation
  - Tests run using the Oracle 11g RDBMS on a quad-core machine
  - Each execution time (in seconds) is the average time of 3 different queries with different selection predicates

Topic hier.	|Group-by|	FT1		FT2	
		Meta-star	Star s.	Meta-star	Star s.
H1	0	13.8	12.7	140.0	137.2
	1	16.0	5.8	174.6	64.3
	2	16.6	14.6	162.4	162.1
H2	0	13.6	13.0	136.0	133.6
	1	16.7	5.6	179.5	179.4
	2	17.0	16.2	175.8	162.2
H3	0	12.2	9.0	139.1	126.6
	1	15.9	14.1	147.3	172.1
	2	35.1	16.9	187.1	144.2

Evaluation

- Performances of Meta-Star are compared with traditional star schemata using queries with Topic Aggregation
  - Tests run using the Oracle 11g RDBMS on a quad-core machine
  - Each execution time (in seconds) is the average time of 3 different queries with different selection predicates

Topic hier.	|Group-by|	FT1		FT2	
		Meta-star	**Star s.**	**Meta-star**	**Star s.**
H1	0	13.8	12.7	140.0	137.2
	1	16.0	5.8	174.6	64.3
	2	16.6	14.6	162.4	162.1
H2	0	13.6	13.0	136.0	133.6
	1	16.7	5.6	179.5	179.4
	2	17.0	16.2	175.8	162.2
H3	0	12.2	9.0	139.1	126.6
	1	15.9	14.1	147.3	172.1
	2	35.1	16.9	187.1	144.2

In most cases star schemata outperform meta-stars, but the gap is quite limited and perfectly acceptable

Evaluation

- Performances of Meta-Star are compared with traditional star schemata using queries with Topic Aggregation
  - Tests run using the Oracle 11g RDBMS on a quad-core machine
  - Each execution time (in seconds) is the average time of 3 different queries with different selection predicates

Topic hier.	|Group-by|	FT1		FT2	
		Meta-star	Star s.	Meta-star	Star s.
H1	0	13.8	12.7	140.0	137.2
	1	16.0	5.8	174.6	64.3
	2	16.6	14.6	162.4	162.1
H2	0	13.6	13.0	136.0	133.6
	1	16.7	5.6	179.5	179.4
	2	17.0	16.2	175.8	162.2
H3	0	12.2	9.0	139.1	126.6
	1	15.9	14.1	147.3	172.1
	2	35.1	16.9	187.1	144.2

The execution time is mostly spent on the fact table, as the increase of execution time is proportional to the increase of the fact table size

Evaluation

- Performances of Meta-Star are compared with traditional star schemata using queries with Topic Aggregation
  - Tests run using the Oracle 11g RDBMS on a quad-core machine
  - Each execution time (in seconds) is the average time of 3 different queries with different selection predicates

Topic hier.	|Group-by|	FT1		FT2	
		Meta-star	Star s.	Meta-star	Star s.
H1	0	13.8	12.7	140.0	137.2
	1	16.0	5.8	174.6	64.3
	2	16.6	14.6	162.4	162.1
H2	0	13.6	13.0	136.0	133.6
	1	16.7	5.6	179.5	179.4
	2	17.0	16.2	175.8	162.2
H3	0	12.2	9.0	139.1	126.6
	1	15.9	14.1	147.3	172.1
	2	35.1	16.9	187.1	144.2

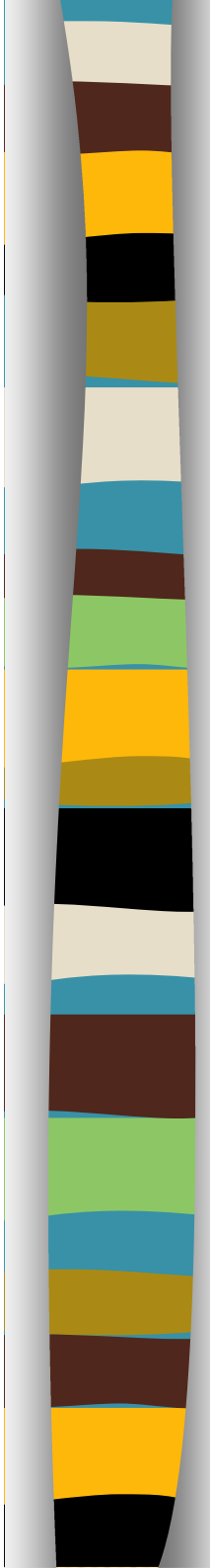
Execution times on the meta-star increase smoothly for group-by's with increasing number of levels

Evaluation

- Performances of Meta-Star are compared with traditional star schemata using queries with Topic Aggregation
  - Tests run using the Oracle 11g RDBMS on a quad-core machine
  - Each execution time (in seconds) is the average time of 3 different queries with different selection predicates

Topic hier.	|Group-by|	FT1		FT2	
		Meta-star	Star s.	Meta-star	Star s.
H1	0	13.8	12.7	140.0	137.2
	1	16.0	5.8	174.6	64.3
	2	16.6	14.6	162.4	162.1
H2	0	13.6	13.0	136.0	133.6
	1	16.7	5.6	179.5	179.4
	2	17.0	16.2	175.8	162.2
H3	0	12.2	9.0	139.1	126.6
	1	15.9	14.1	147.3	172.1
	2	35.1	16.9	187.1	144.2

Execution times on the meta-star increase slowly for topic and roll-up tables with increasing cardinality



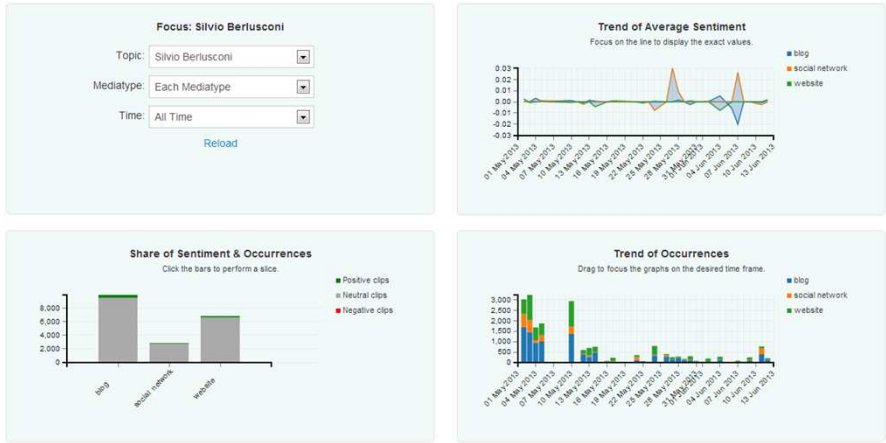
Conclusions

- Meta-Stars: an expressive solution to model topic hierarchies for SBI
  - We proposed an approach that handles dynamics of topic classification, integrability with business hierarchies, semantics-aware queries
  - We analyzed its performance against traditional star schemata

Conclusions

- Meta-Stars: an expressive solution to model topic hierarchies for SBI
  - We proposed an approach that handles dynamics of topic classification, integrability with business hierarchies, semantics-aware queries
  - We analyzed its performance against traditional star schemata

- A real-life experiment is being conducted:
  - **http://semantic.csr.unibo.it/**
  - Implemented a meta-star with a topic hierarchy on Italian politics
  - User interface allows for simple OLAP analysis



Conclusions

- Now working on:
  - Cost model for Meta-Stars
    Studying functions that allow the size and efficiency of a topic hierarchy to be evaluated a priori
  - Topic Hierarchy generation
    Developing an automatic task to turn a topic hierarchy modeled as an ontology into a Meta-Star
  - Coupling SQL and OWL
    Study the possibility of using the OWL language to directly query the topic hierarchy
  - Summarizability for N-M relationships
    Study which summarization rationales are valid and can be adopted to produce interesting results
  - Historicization
    Extending the architecture to handle topic and roll-up semantics variations over time
  - OLAP front-end
    Investigate how commercial OLAP front-ends can be extended to efficiently support meta-stars