A Role of Ontology in Social Data Analytics

Sam Oh, Professor, SKKU iSchool
Seonghun Kim, Post-Doc, SKKU iSchool
Jinhyeong Lee, Director, LGcns SmartSMA

2015. 12. 9

SUNG KYUN KWAN UNIVERSITY

Talk Outline

1. Research Background
2. Research Purpose
3. Research Process
4. Research Findings
5. Further Study

1. Research Background

- Vocabulary Management for Social Data Analytics
  - Tends to be managed in a solo
  - Lack of reusability and interoperability from one app to the other
  - Speedy vocabulary management is needed for live data analytics
- Much efforts and time spent in collecting terms and devising a structure to them
  - Requires a considerable time to establish a VOC management system for a new project
- Relationships among terms and additional details on each term need to be established
  - Useful in deriving insights by taking the relationships into account

I. Research Background

- Relationships among terms and additional details on each term need to be established
  - Useful in deriving insights by taking the relationships into account

Collecting social data

Analyzing Lexicons

Analyzing Lexical Frequencies

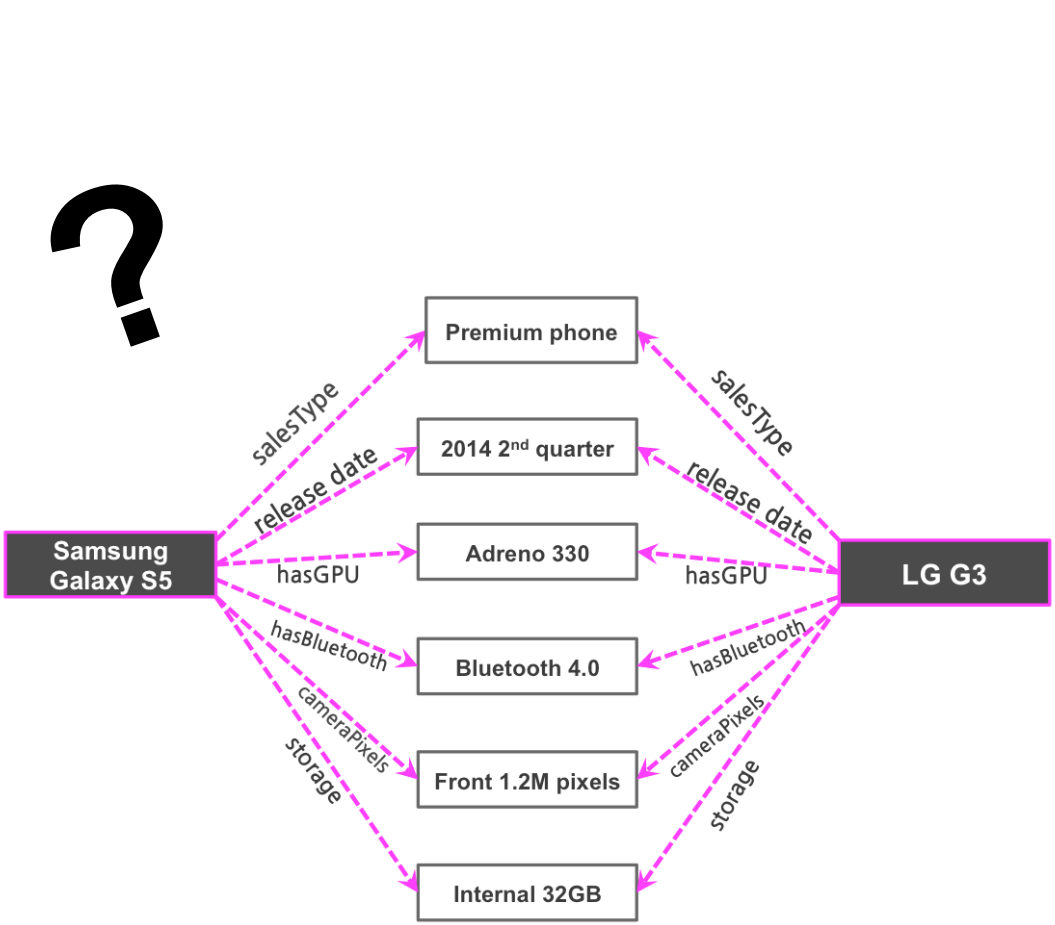
Setting Analysis Framework

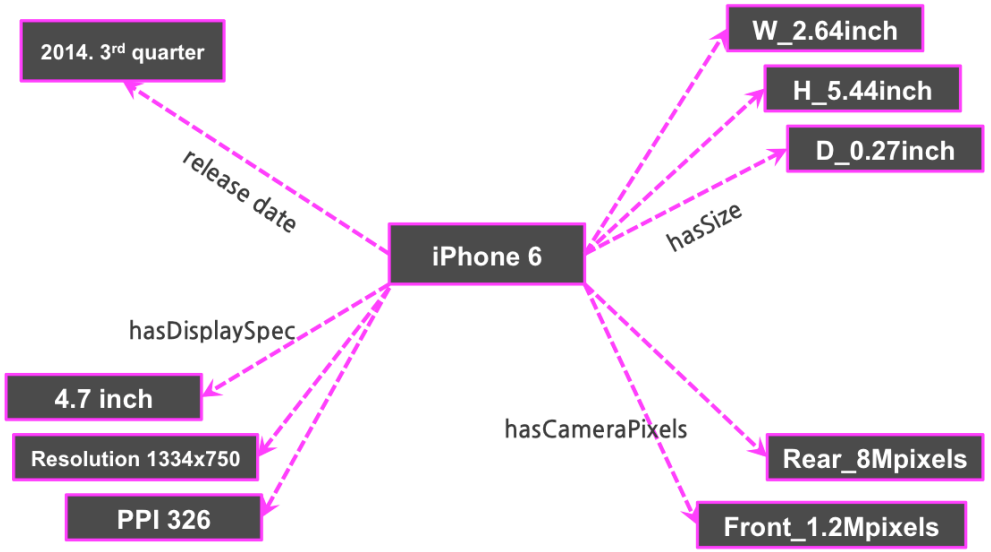
Conducting Classification

Analyzing Classification Results

2. Research Purpose

- To understand what roles ontology can play in social data analytics





3. Research Process

3.1 Constructing 'smartphone' ontology

3.2 Ingesting DBPedia data related to smartphones

3.3 Visualizing 'smartphone' ontology

3.4 Ontology Benefits in Social Data Analytics (SDA)

- Limitations of Current SDA Systems
- Improvement of Current SDA Systems

3.1 Constructing a 'Smartphone' Ontology

Managing smartphone ontology with the following SKOS and OWL vocabularies: skos:Concept, skos:broader, skos:narrower, skos:prefLabel, skos:altLabel

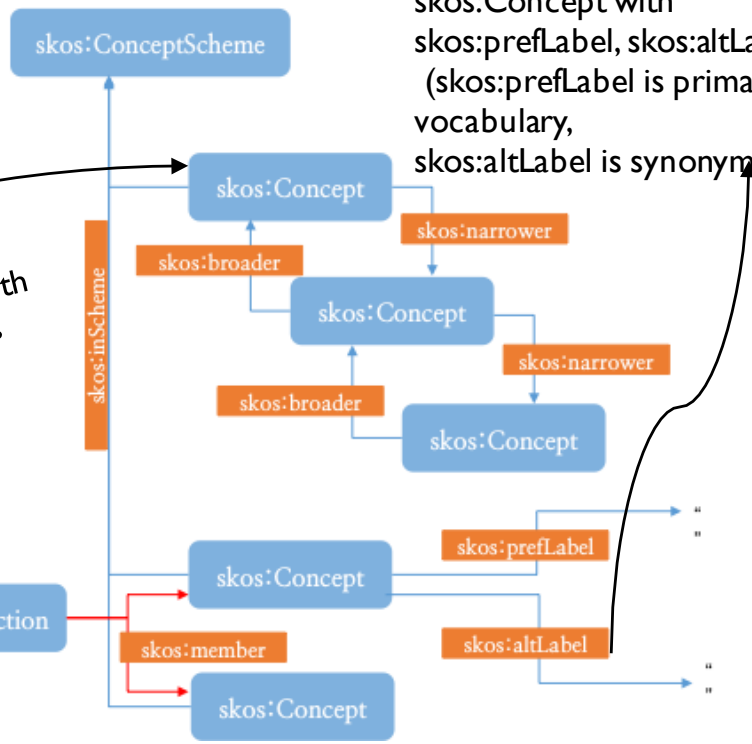
<LG Vocabulary Scheme>

L1	L2	L3	L4	L5
LG Electronics	Smart Phone	Optimus	general	OptimusLTE 2
				OptimusLTE 3
		G Series	premium	G3
				G2
				G PRO
		Nexus	general	Nexus 4
Samsung Electronics	Smart Phone	Galaxy	general	Galaxy Grand
				Galaxy POP
			premium	Galaxy Note3
				Galaxy S4
				Galaxy S5
Apple	Smart Phone	iPhone	premium	iPhone5
				iPhone5s
				iPhone6
				iPhone6plus
			general	iPhone5c

Managing Smartphone Title with skos:Concept to set relations skos:broader, skos:narrower

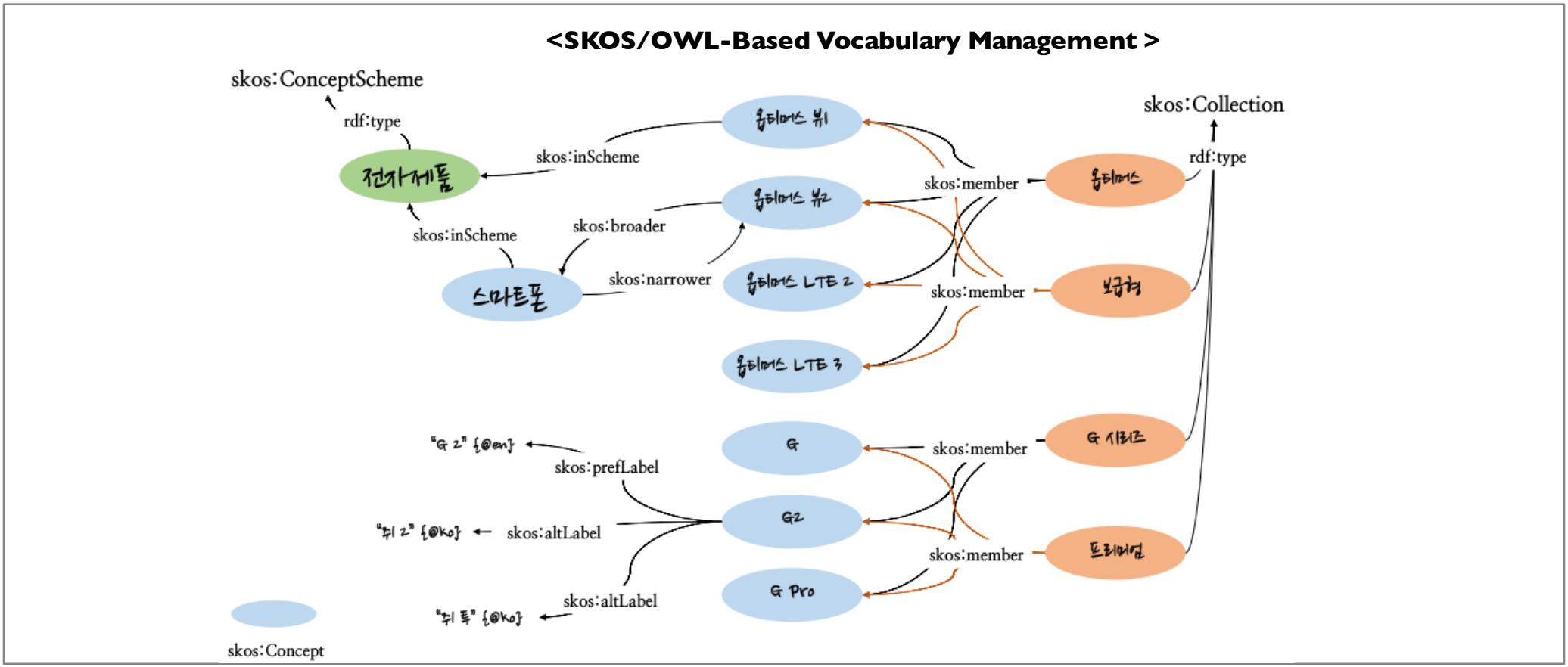
Providing group information using skos:Collection

<SKOS Basic Structure>



Providing synonym for skos:Concept with skos:prefLabel, skos:altLabel (skos:prefLabel is primary vocabulary, skos:altLabel is synonym)

3.1 Constructing a 'Smartphone' Ontology



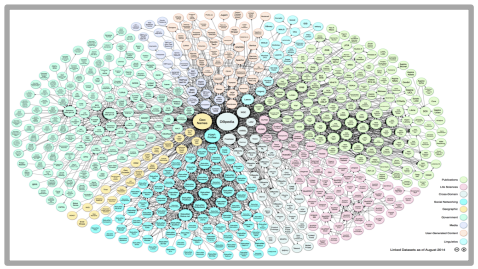
SKOS/OWL-based Management

- **SKOS/OWL-based vocabulary management allows LG-CNS to import external data and provides interoperability.**

3.2 Ingesting DBPedia Data

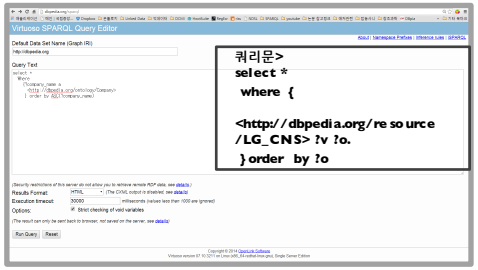
Ingesting relevant data (firms and personal info) into ontology

DBpedia



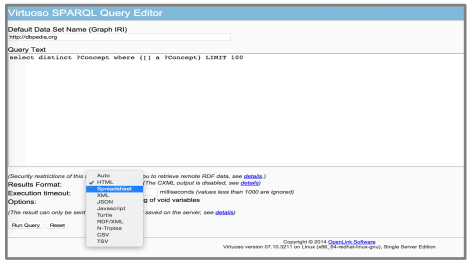
- As of August 2014, 500 datasets. The center circle is Dbpedia.
- LOD De facto, Huge amount of data
- 191 Korean firm information out of 50,000 firm information.
- 200 Korean out of 1,440,000 personal information

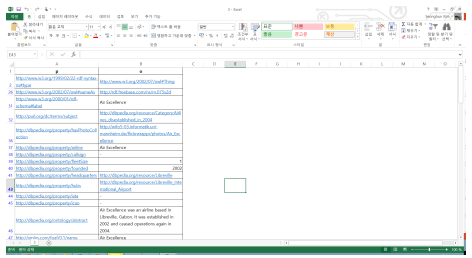
Querying Data



Extract metadata for massive data extracted from Dbpedia

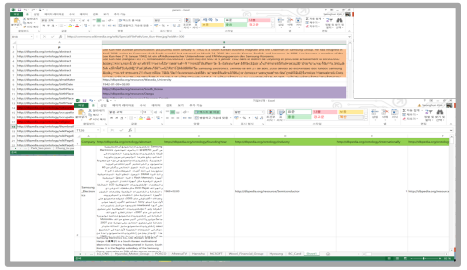
Downloading Data





SPARQL Endpoint – save as spreadsheet

Saving as a spreadsheet



Have a program to import spreadsheet data into ontology

3.3 Visualizing 'Smartphone' Ontology

TM Analysis Expanded

- Protégé PlugIn, Relfinder, and D3 provide different views of data

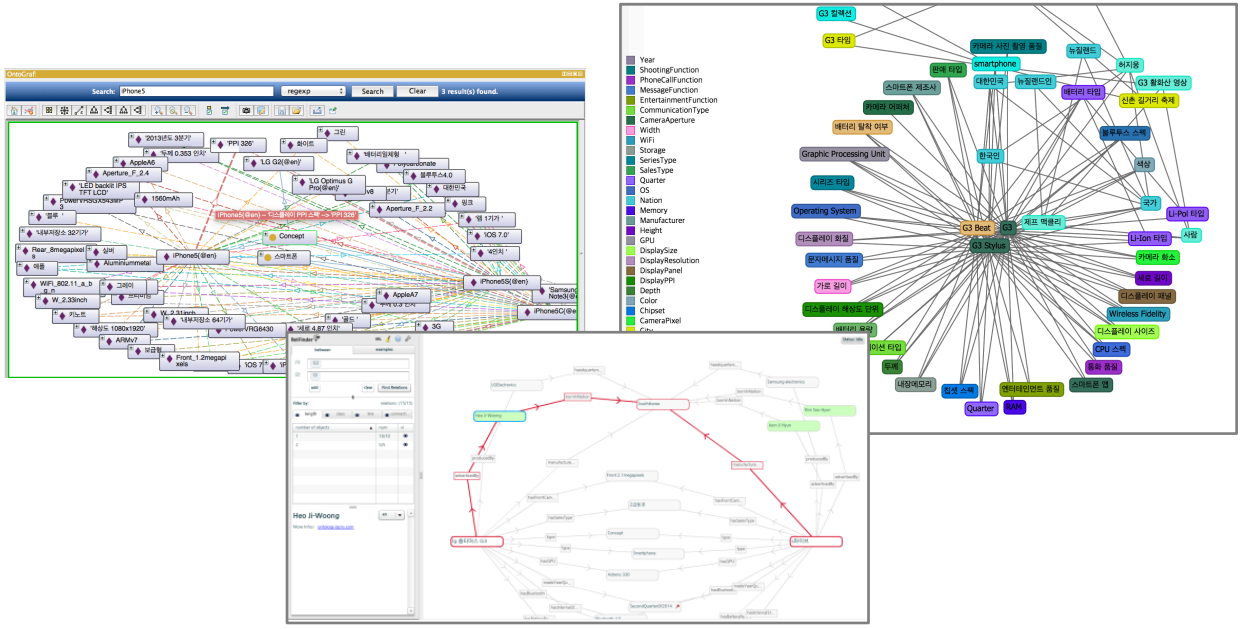
AS-IS

- Proximity measured by word-frequency
- Raw data must be consulted to understand relationships among terms

TO-BE

- Enhanced views of data
- Relationships among keywords are provided

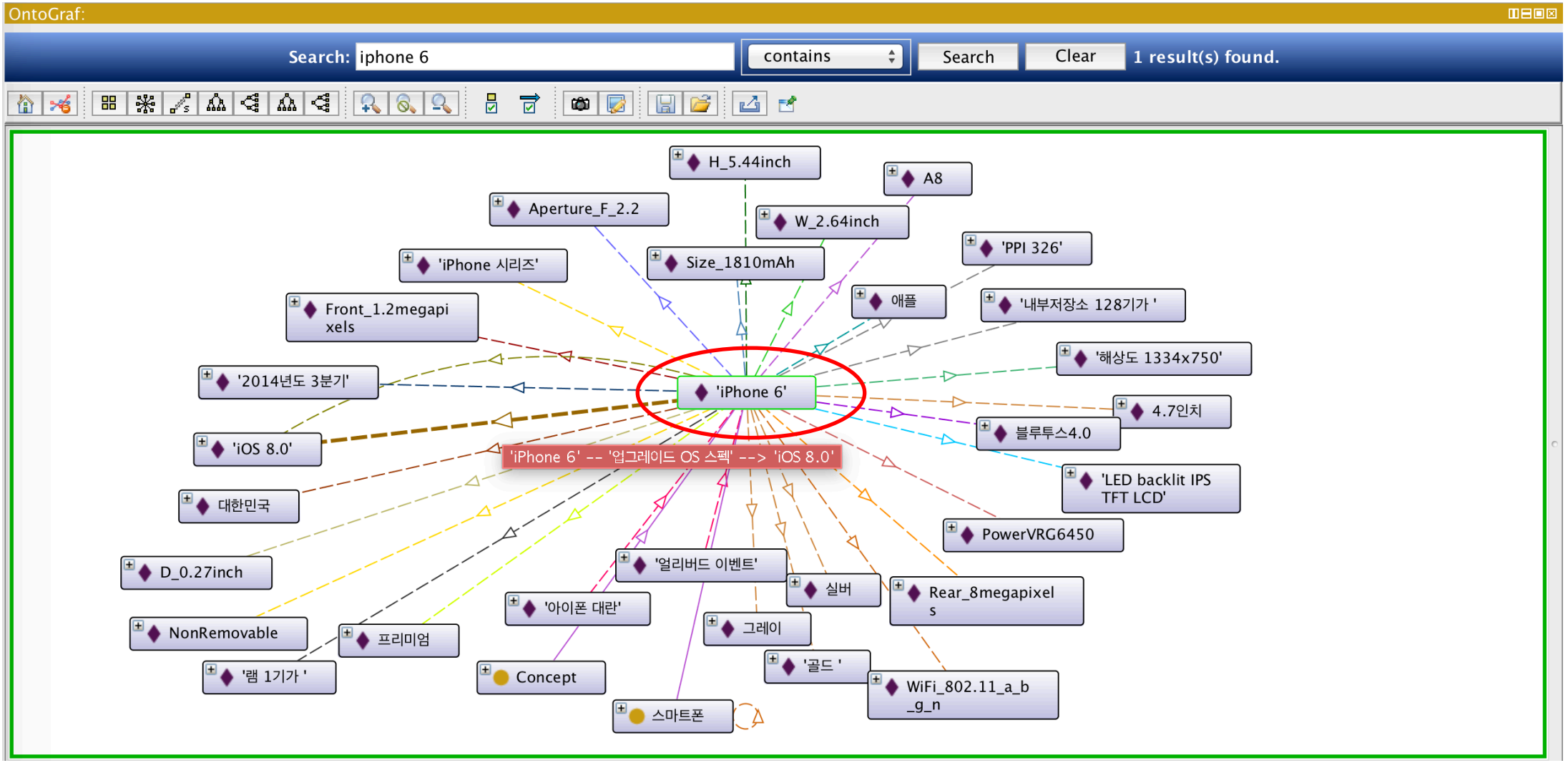
AS-IS TM



3.3 Visualizing 'Smartphone' Ontology: Type A

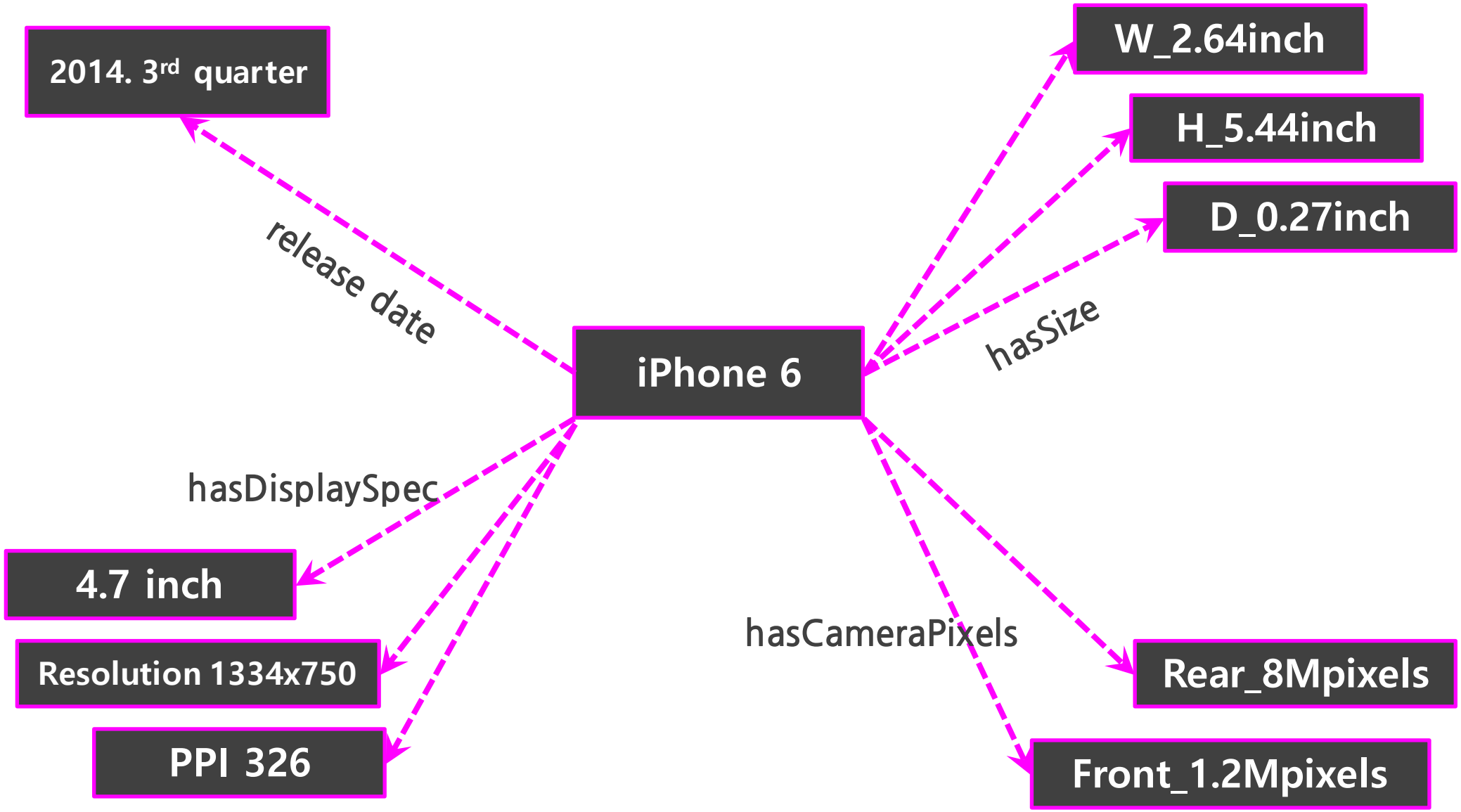
Main features of Type A

- **Providing detailed info on each term**



Mouse-over will display relationship

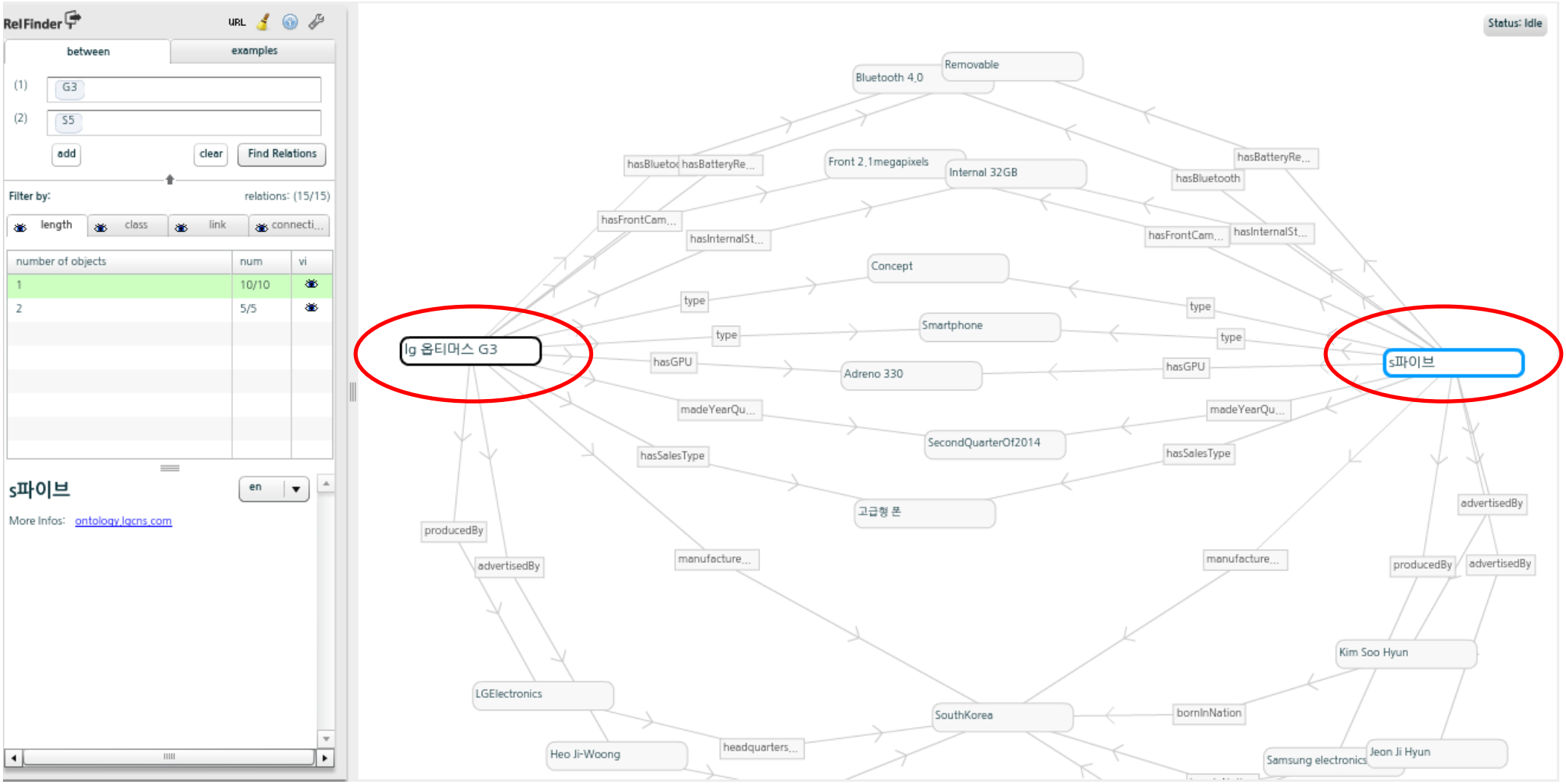
Main features of Type A



3.3 Visualizing ‘Smartphone’ Ontology: Type B

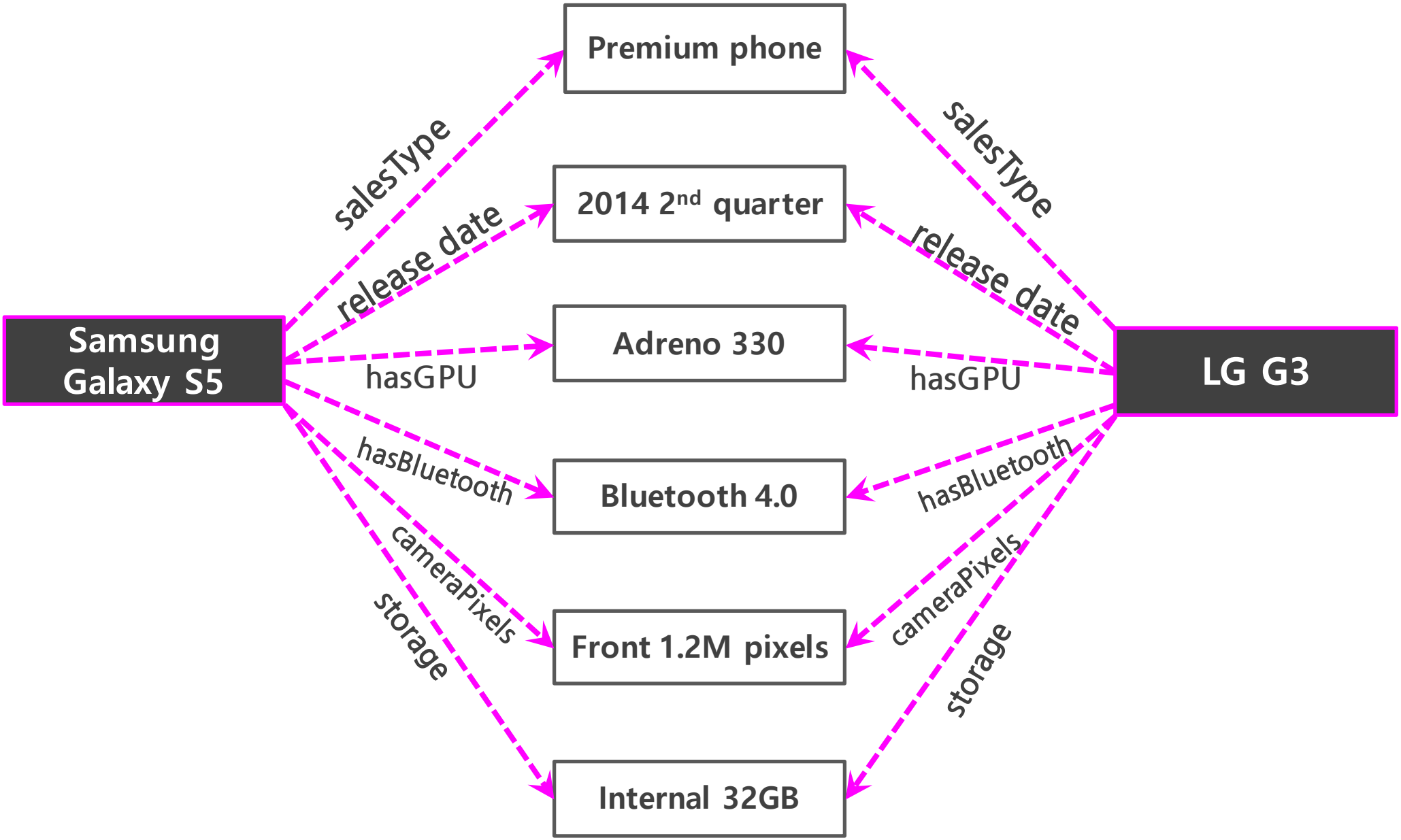
Main features of Type B

- **When you want to know relationships between two terms**



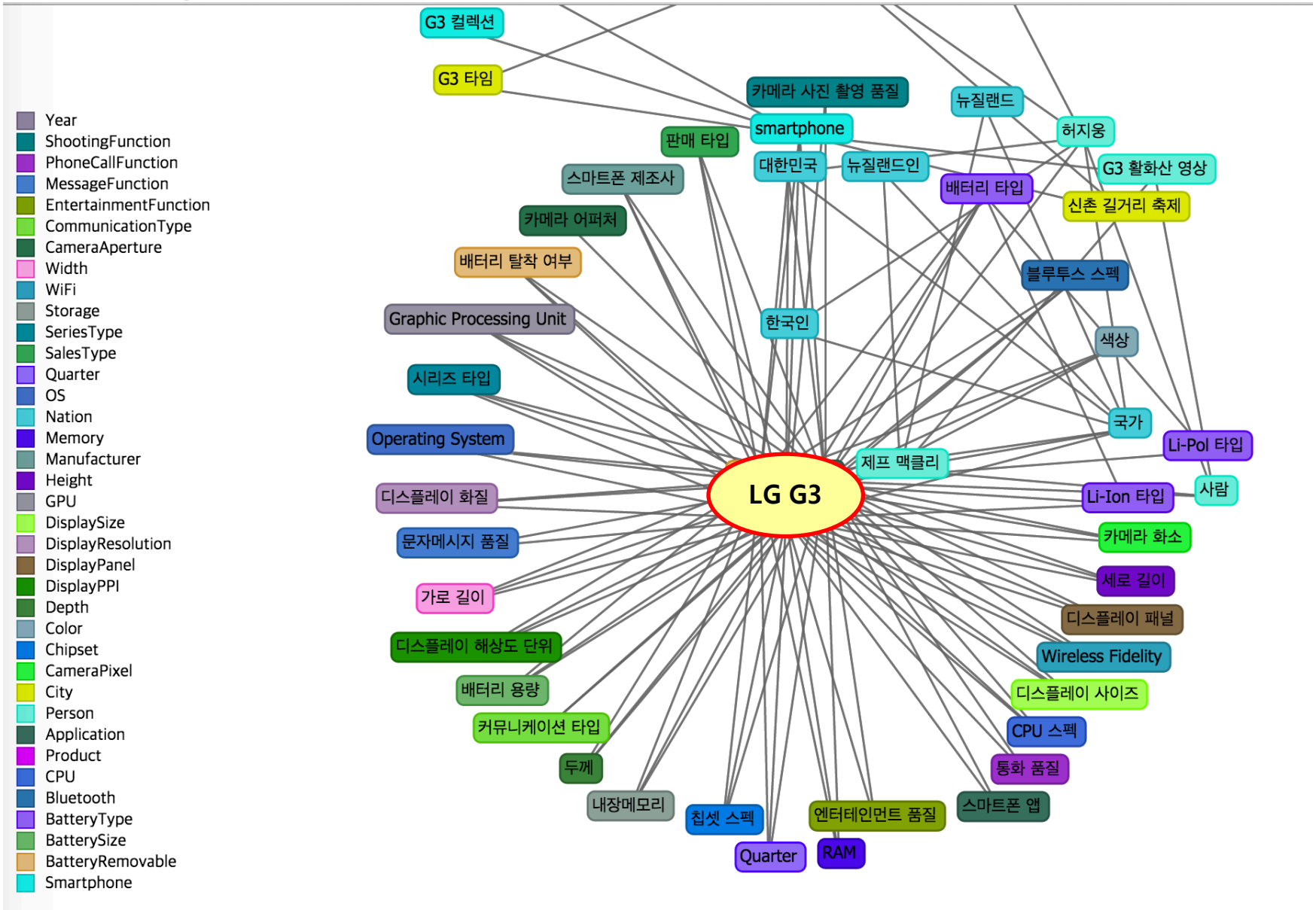
- **For example, what are the common features of G3 and S5?**

Main features of Type B



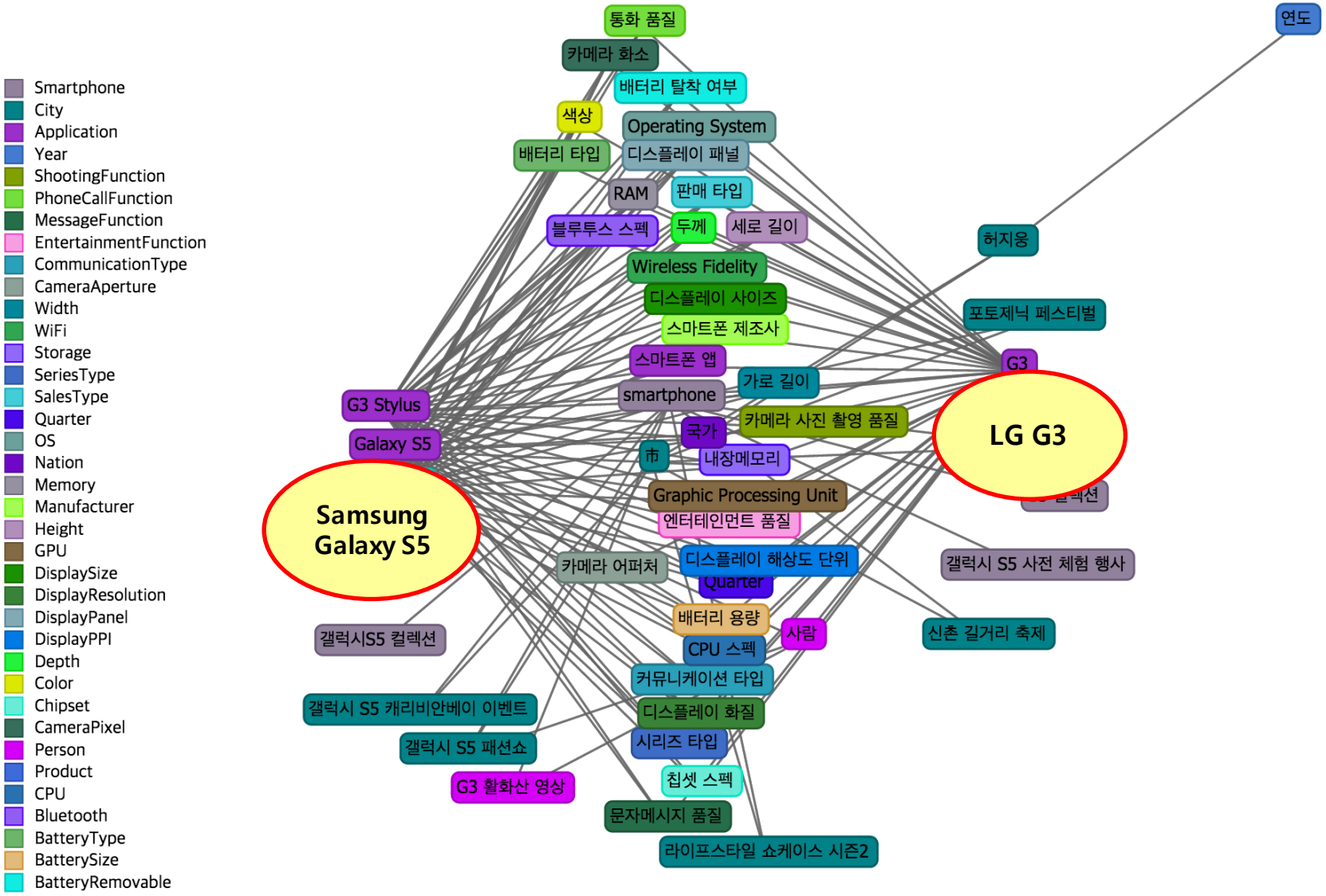
3.3 Visualizing 'Smartphone' Ontology: Type C

Combining Type A and B



Search Results of 'LG G3'

3.3 Visualizing ‘Smartphone’ Ontology: Type C



Search Results of ‘LG G3’ and ‘Samsung Galaxy S5’

3.4 Ontology Benefits in Social Data Analytics

To identify the benefits of ontology in social data analytics

- Conducted Delphi surveys involving experts in social data analytics and social marketing
- Extracted criteria to assess the benefits of the ontology in social data analytics

3.4 A Current Social Data Analysis (SDA) System

Experts in Social Data Analytics

- ✓Those who are savvy in setting up rules and stop words to take advantage of a particular domain analysis
- ✓Those who are savvy in building a comprehensive service such as a TM dashboard
- ✓Those who have worked for more than 2 years in the job mentioned above

Experts in Social Marketing

- ✓Those who have expertise in analyzing markets and developing an effective marketing plan utilizing SDA tools
- ✓Those who have experience in writing 'insight report' on market and product analysis

3.4 Ontology Benefits in Social Data Analytics

1st Delphi Study

Questions used

- What are the benefits of the current SAS Text Mining (TM) tools in conducting SDA?
- What are the shortcomings of the current SAS TM tools in conducting SDA?
- To have a more effective SAS TM tools in SDA, what functions should be improved

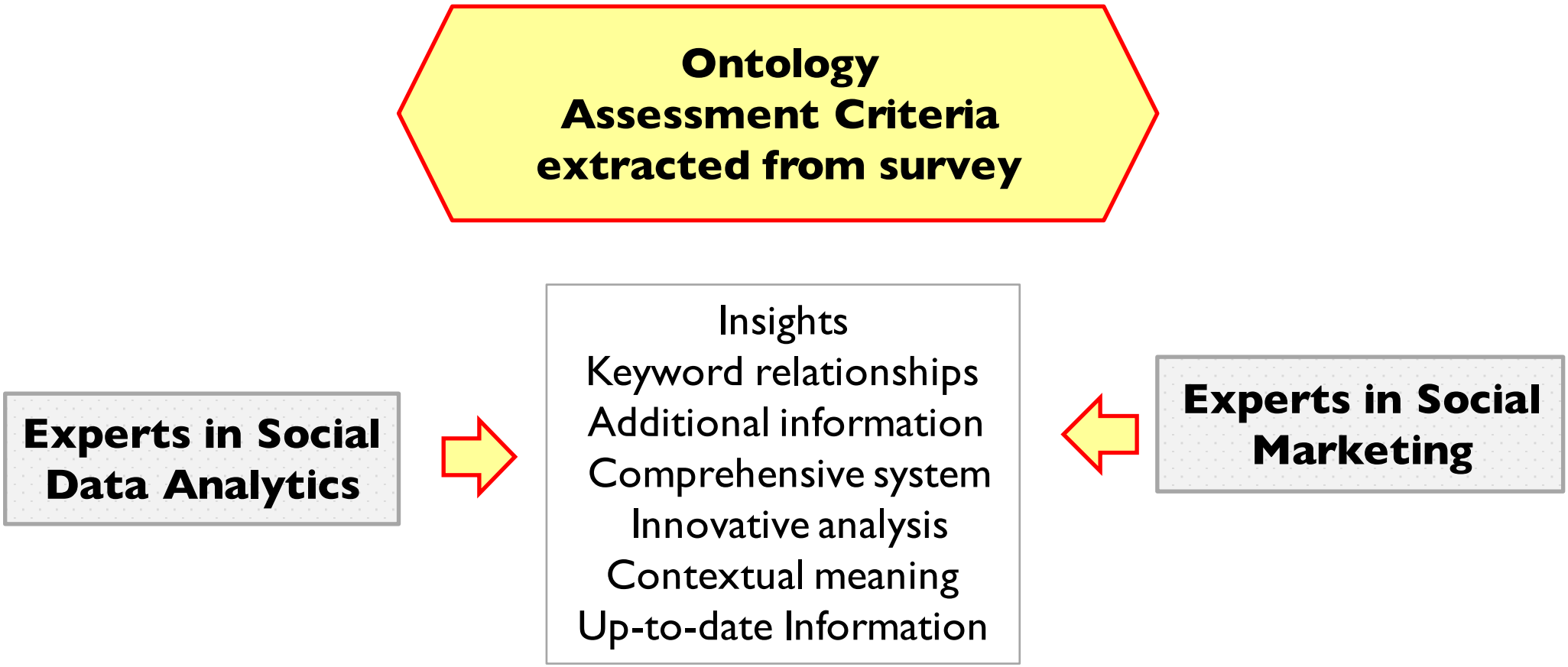
3.4 Ontology Benefits in Social Data Analytics

Quantitative Evaluation

Questions used for study

- Was it helpful to use ontology in deriving insights?
- In understanding keyword relationships
- In obtaining further details on each term
- In constructing a comprehensive system for social data analytics
- In combining rule-based analysis and machine learning
- In getting contextual meaning of each term in a particular domain
- In obtaining up-to-date information

3.4 Ontology Benefits in Social Data Analytics



3.4 Ontology Benefits in Social Data Analytics

Results of Quantitative Evaluation: Experts in Social Data Analytics

- **7 Point Scale**

Variable	Avg	STD	Normal	T-test p-v	Wilcoxon p-v
Insights	5.286	0.731	0.712	0.341	0.303
Keyword relationships	5.286	0.591	0.744	0.248	0.221
Additional info	5.048	0.591	0.450	0.838	0.785
Comprehensive system	5.333	1.054	0.952	0.435	0.336
Innovative analyses	5.238	1.384	0998	0.665	0.598
Contextual meaning	5.333	0.923	0.910	0.376	0.336
Up-to-date info	4.476	1.425	0.866	0.368	0.380

3.4 Ontology Benefits in Social Data Analytics

Results of Quantitative Evaluation: Experts in Social Marketing

- **7 Point Scale**

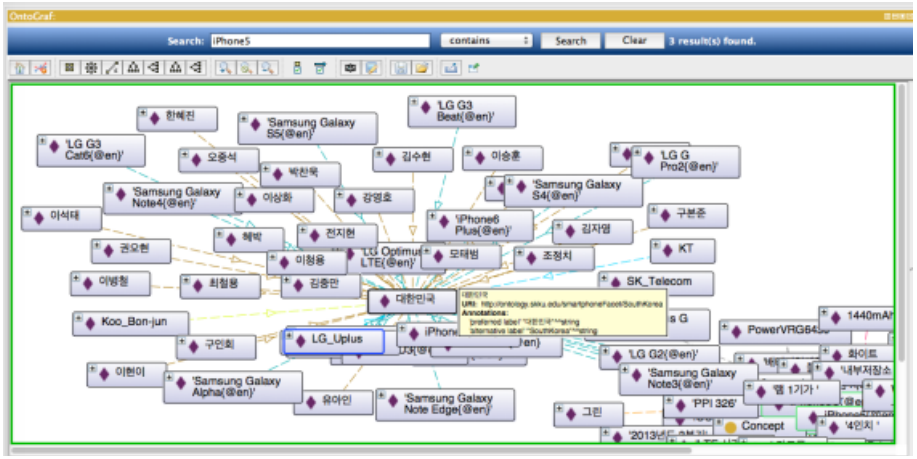
Variable	Avg	STD	Normal	T-test p-v	Wilcoxon p-v
Insights	5.067	0.723	0.925	0.847	0.892
Keyword relationships	5.067	0.760	0.949	0.854	0.705
Additional info	5.533	0.606	0.955	0.120	0.109
Comprehensive system	5.333	0408	0.759	0.142	0.102
Innovative analyses	4.733	1.140	0.993	0.629	0.715
Contextual meaning	4.800	0.960	0.990	0.666	0.581
Up-to-date info	4.476	1.193	0.908	0.374	0.336

3.4 Ontology Benefits in Social Data Analytics

Qualitative Evaluation I

- We asked why they prefer a particular visualization over the others per question (only those Qs that demonstrated significant difference)?
- Experts in Social Data Analytics
  - Insights, Keyword relationships, Innovative analyses (3)
- Experts in Social Marketing
  - Insights, Comprehensive systems (2)

Type A: Additional Info



Type B: Relationship

Qualitative Evaluation 2

- Further analysis on deriving insights
  - Experts in Social Data Analytics: Preferred Type B
    - Readability, Identifying Relationships of each term
  - Experts in Social Marketing: Preferred Type A
    - Additional detail info was more important to them in gaining insights
    - Identifying relationships was optional, not crucial to them

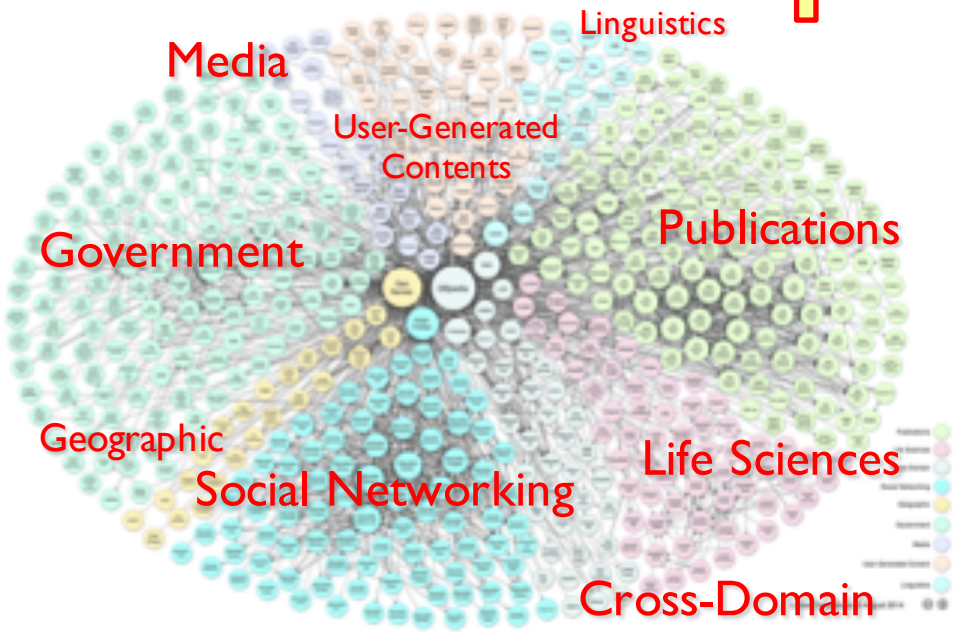
Qualitative Evaluation 3

What are additional roles that ontology can play in SDA?

Additional Roles of Ontology

- Experts in Social Market did not submit any idea on this.
- Experts in Social Data Analytics suggested that the following features are desired from the beginning of ontology setup.
  - Definition
  - Synonyms
  - Relationships

4. Further Study



Datasets by topical domain.

Topic	Datasets	%
Government	183	18.05%
Publications	96	9.47%
Life sciences	83	8.19%
User-generated content	48	4.73%
Cross-domain	41	4.04%
Media	22	2.17%
Geographic	21	2.07%
Social web	520	51.28%
Total	1014	

- Global ID Management
- Ways to integrate LOD into ontology to enhance SDA

Thank you for your attention!!
Any questions?