Internet Review Opinion Mining utilizing Opinion Mining and Data Visualization

Seongmin Mun [*] , Ginam Kim [†] , Raja Mubashar Karim [‡] , Kyungwon Lee [§]

[*†‡] Life media Interdisciplinary Programs [§] Department of Digital Media, Ajou University

[*]skdkflak@naver.com, [†]mocca002@naver.com, [‡]mubashiry978@gmai.com, [§]kwlee@ajou.ac.kr
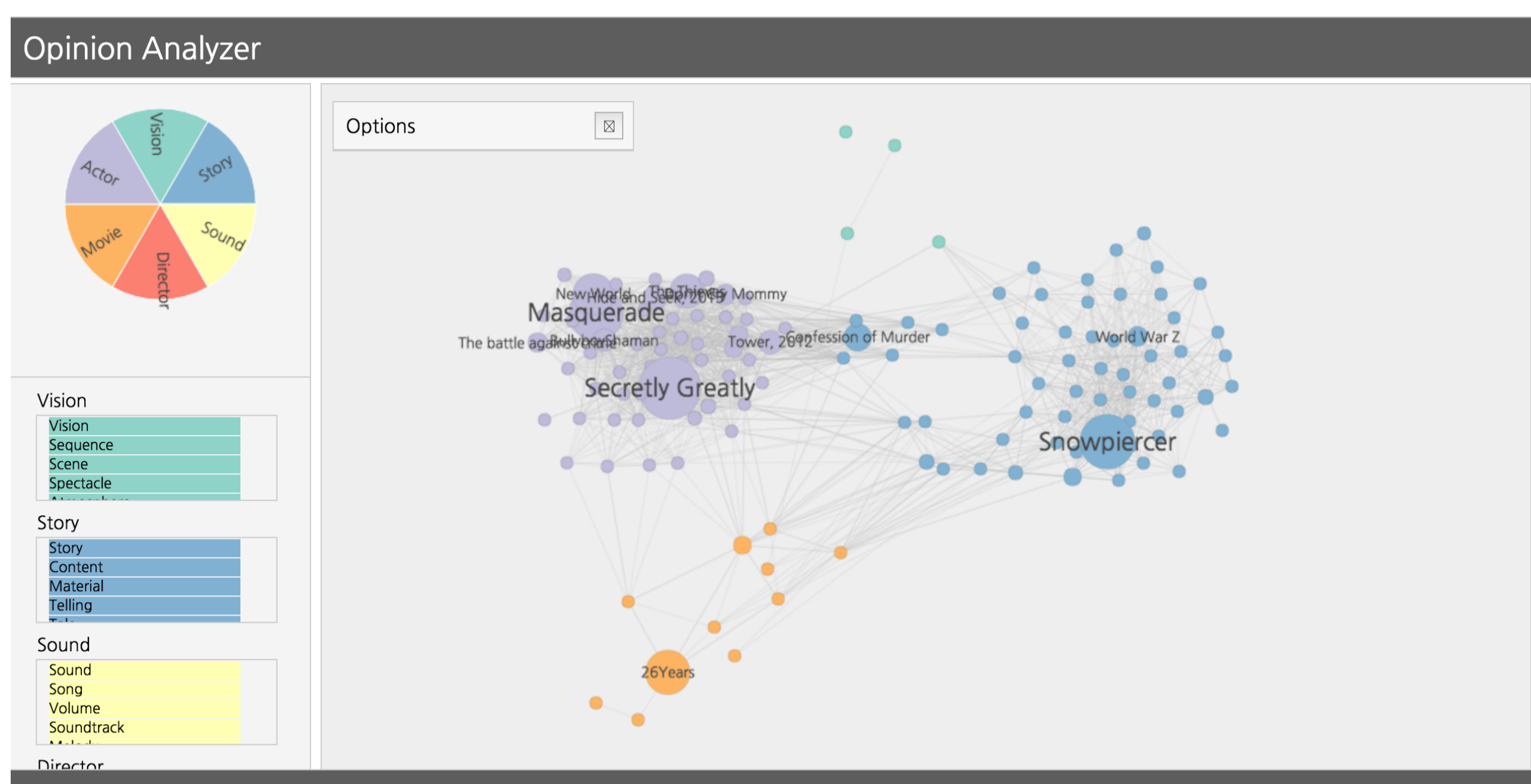
Figure 1.   Over view (Demo site URL : hellomocca.github.io/demo/)
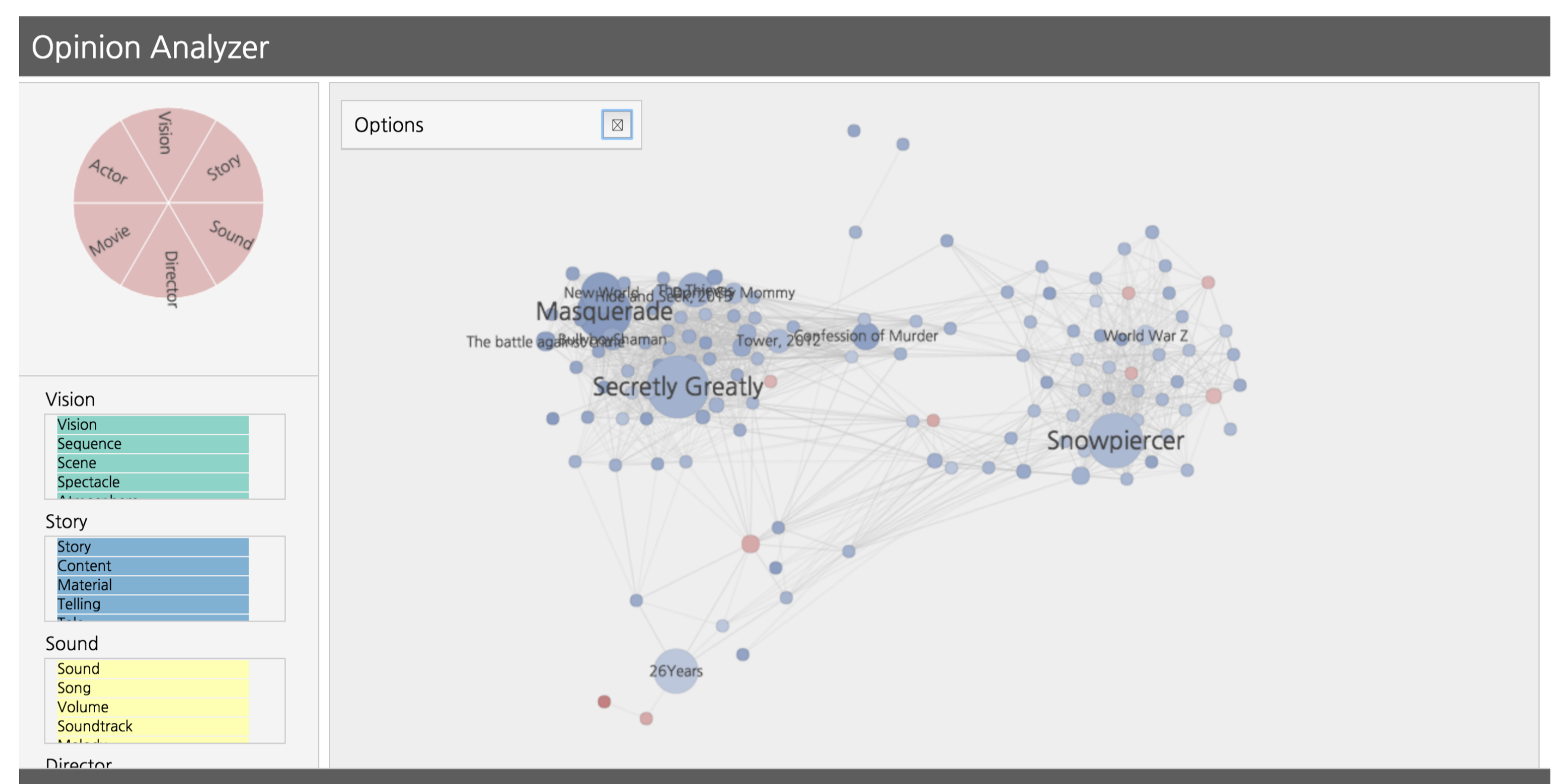

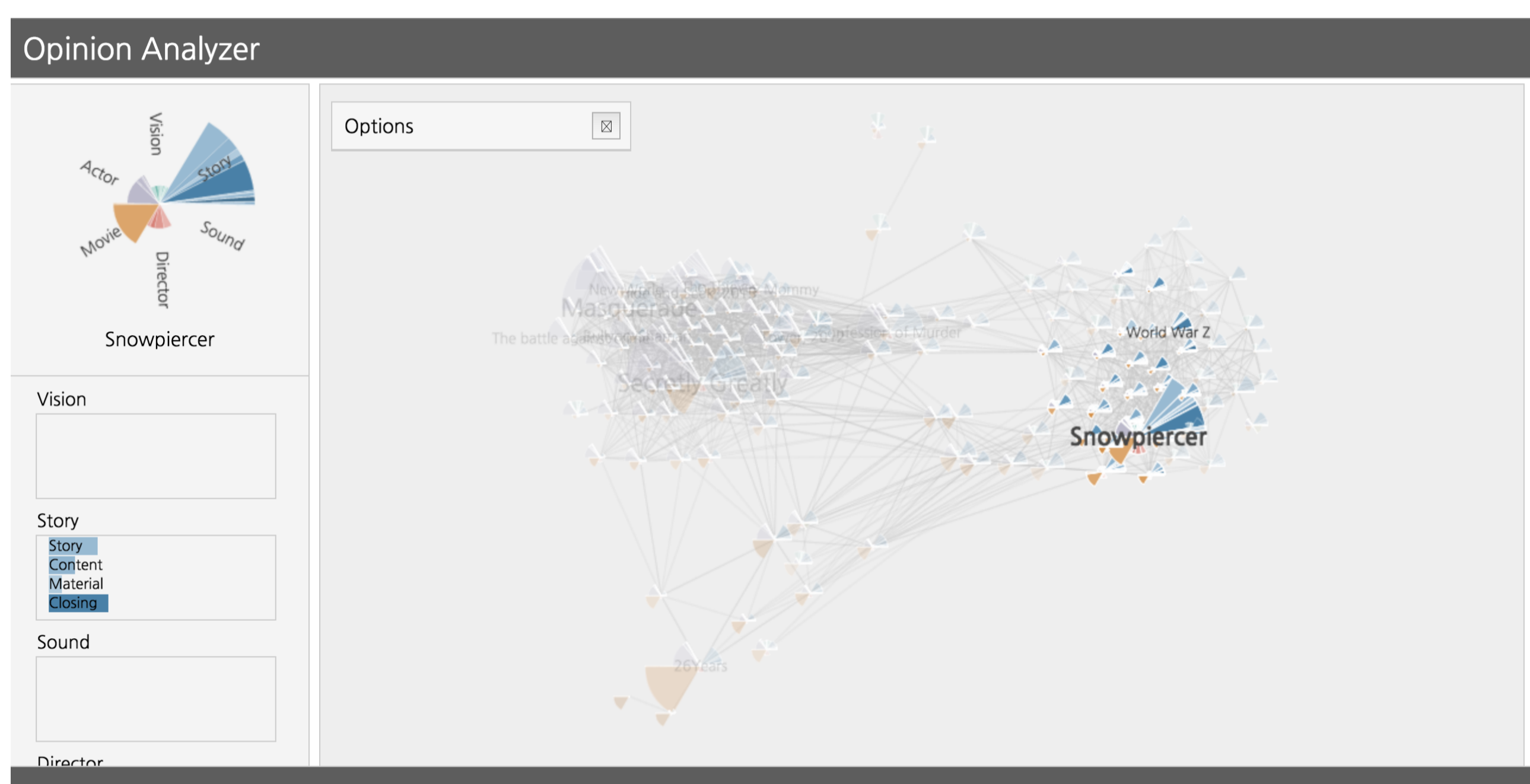Figure 2.   Over view colored by Polarity (Positive : Blue, Negative : Red)

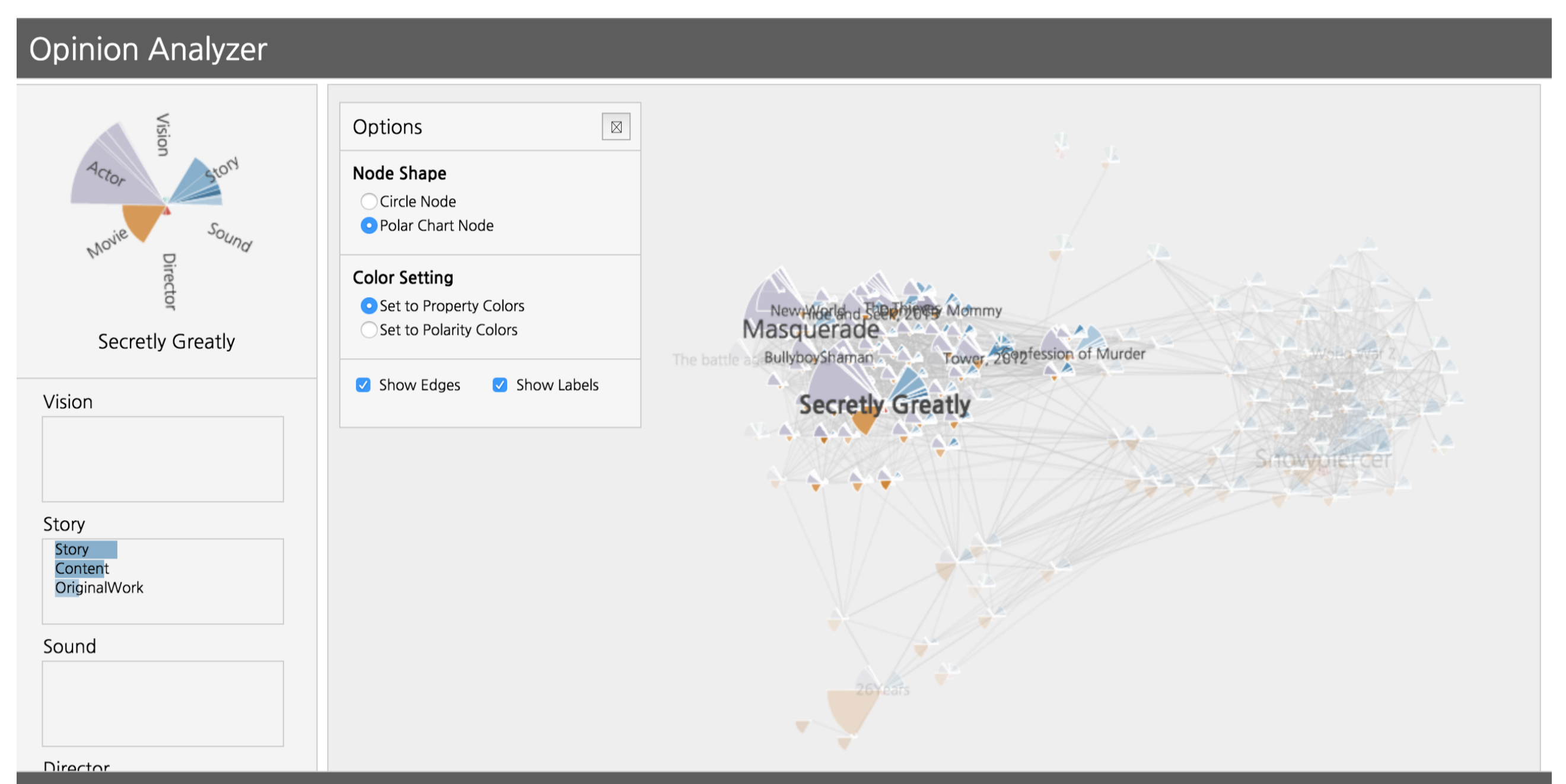Figure 3.   The clustering group centered by "Snowpiercer"

Figure 4.   Detail view when users click on a node named "Secretly Greatly"

Introduction

The research is conducting about the effect of online information to society by producing online information from internet user after the web 2.0age. However, there is limitation that it costs a lot to analyze online information by human and to solve this, opinion minding is used which is automation system to analyze online information. This research constructs ontology of limited domain by "movie" among many reviews which we can find on the internet, and suggests ontology and visualized opinion mining base on what we suggested.

Constructing Ontology

Ontology derived from Greek ontos which means "presence", and logos which means "word", we can define that ontology is proposedly standardized statement to conceptualize some specific domain. First, we need to pick category of ontology class and Property evaluation element keyword which is related with each category to construct ontology for opinion mining[1]. In this research, we abstract index from review data with category of Li Zhuang movie element, and categorize and pick keyword.

```
<Class: Character>              (1)
    <property: characterization>
        <value: 123>
    <property: role>
        <value: 234>
```

Evaluation Vocabulary

In this research, we tried to improve efficiency of opinion minding procedure by constructing special predicate dictionary according to category of subject of evaluation element. For example, "scenario is simple." this sentence is labeled as following in the procedure of morpheme analysis[2].

```
Class    Pdt    TagPdt    Pol      (2)
```

This sentence is treated meaningful sentence because it contains "scenario" keyword which include in "story" category.[3] And this sentence, predicate "simple" is showed up for "scenario", we add this in predicate dictionary about "story". Class means category of relevant predicate, it is selected relevant predicate as search object when keyword is in the sentence which is included in category. This class becomes "story" in higher example sentence. Pdt is original form of predicate in the body. It becomes "simple" in higher example sentence. TagPdt means predicate which morpheme is separated. Predicate is also needed separated morpheme form because we use sentence which is separated in morpheme in analyzing step. It becomes "단순/NNG 하/XSV" in higher example sentence. Pol means polarity which predicate has, it signifies 1 in positive case, -1 in negative case by reference from prior research.
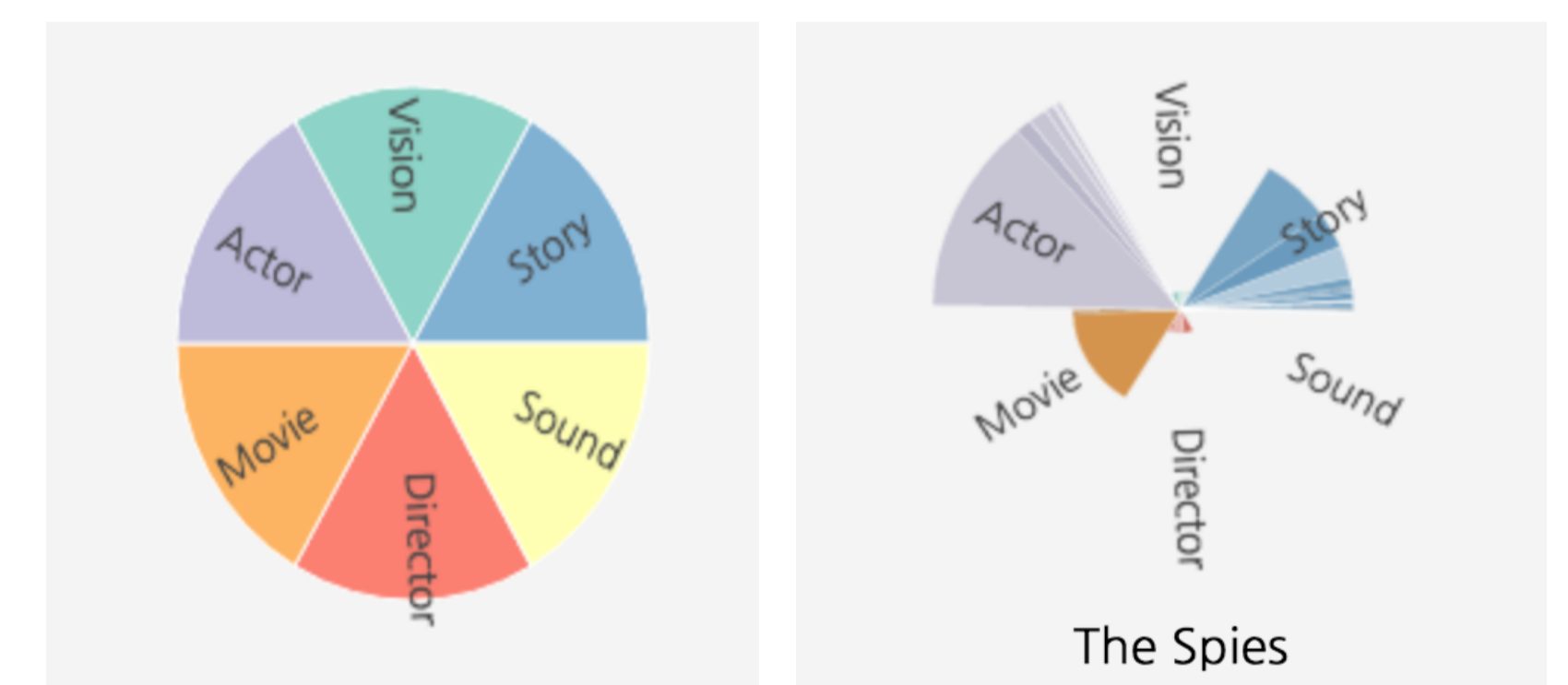
$$시나리오/NNG \ 가/JKS \ 단순/NNG \ 하/XSV \ 다/EFN \ ./SF \quad (3)$$

Ontology Visualization

This research chooses 130 movies which have more than 8000 review from 2011 to 2013. We analyze and visualize whole audiences' opinion through opinion minding. The function of opinion visualization is as following. First is visualization for each node. Circle graph shows frequency estimation and intense for each evaluation element toward 6 ways which are 12 o'clock, 2 o'clock, 4 o'clock, 6 o'clock, 8 o'clock, and 10 o'clock. We can see through graph that there are many evaluation about actor in the example of "Secretly Greatly" We can check the result of specific evaluation element in bar graph. Second is group visualization. This research is calculated by the result of opinion minding to divide into group by product which has similar valuation. Vector value of each product to calculate similarity is ratio per evaluation element among all the review and calculated by cosine similarity. Cosine similarity can be resulted as following when we have 'n' unit of vector 'A' and 'B'[4].

$$similarity = \cos(\theta) = \frac{A \cdot B}{\| A \| \ \| B \|} = \frac{\sum_{i=1}^{n} A_i \times B_i}{\sqrt{\sum_{i=1}^{n}(A_i)^2 \times (B_i)^2}} \quad (4)$$

In this case study, final built group could be divided into three groups. Third is visualization of network. If we use network, we can check that node which is situated between group and group has similar feature with which data. For example, following picture express as a link "Enemies In-Law" in center of group with similar movie. y trend in a chronological order.

The Spies

Conclusion

Significance and result of this research is as following. First, we checked that audiences' overall opinions are different by visualization analysis to each movie, and 130 movies which are used to analyze by audience's opinion are divided into 3 groups. We also deducted that we can check similar movie group with each movie by using grouping visualization. Second, most researches were analyzing data by using ontology which is used normally in original research. Most researches also were used only noun category in the case of opinion minding. However, in our research, we construct ontology of movie domain by using keyword extract and topic remodeling not by using original ontology. It is also meaningful that we applied way of understanding of argument structure in opinion mining with noun category.

References

[1] Hui Yang, Minjie Zhang. Ontology-based resource descriptions for distributed information sources. Information Technology and Applications, 2005. ICITA 2005. Third International Conference on, pages 143-148. IEEE Computer Society. 2005.

[2] Chao He, Yu-feng Zhang. Research on semantic association pattern mining model based on ontology. Advanced Computer Theory and Engineering (ICACTE), 2010 3rd International Conference on, volume 1, pages 497-501. IEEE Computer Society, 2010.