

Voice of the Customers

: Mining Online Customer Reviews for Product Feature-Based Ranking  
의 리뷰

# KING BLABLA AND IDLES

Presented by : kING ash dragon

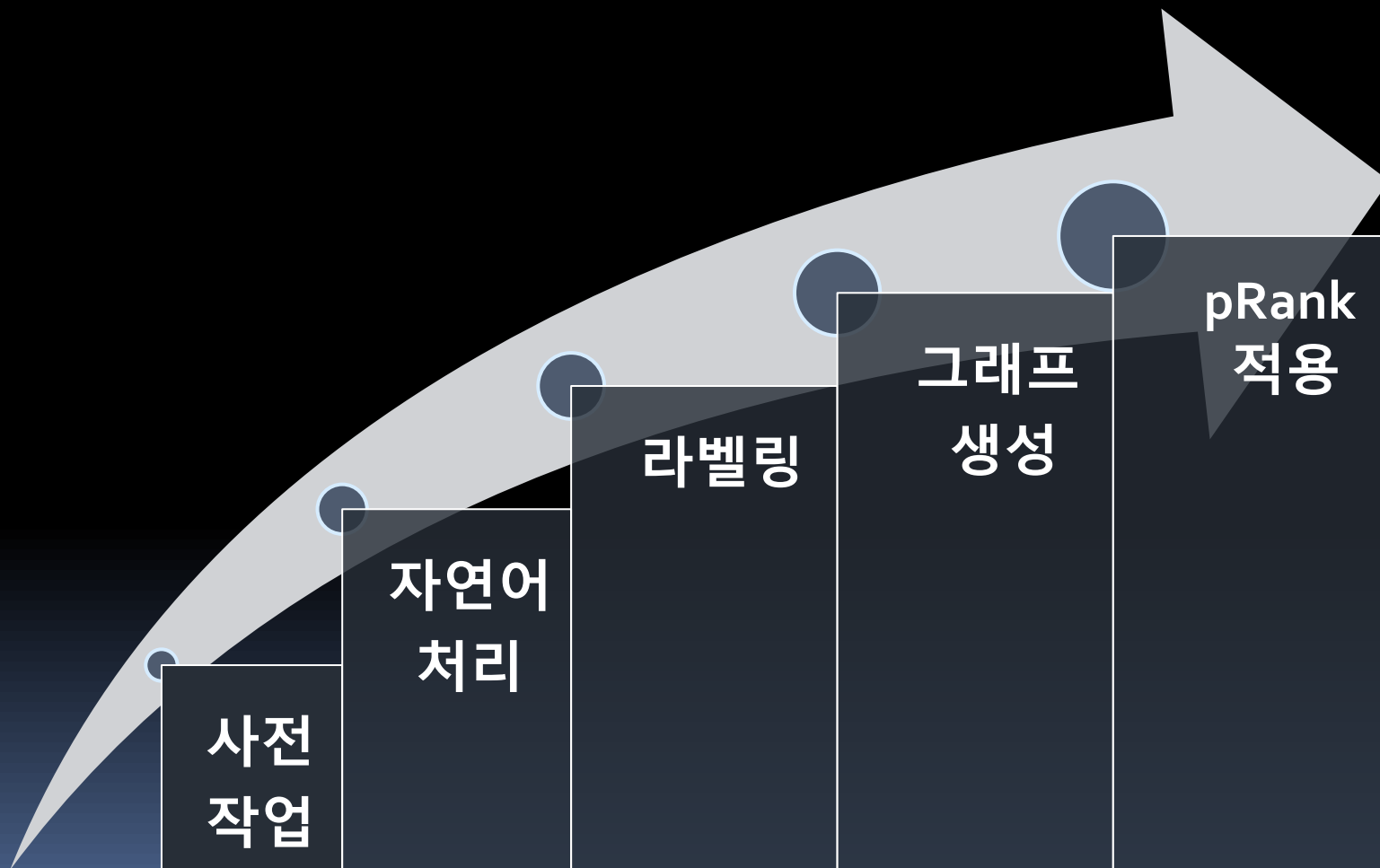
고원희 / 심재용 / 유영창 / 임진하

# 논문의 목적

- 신뢰도 : 판매자가 제공 정보 < '타 고객의 리뷰'
- 각 기능에 따라 소비자의 선호도가 다름

➔ 즉, 리뷰를 기반으로 기능(feature)에 따른  
랭킹 산출이 목적!

# 논문의 개괄적 흐름



# 사전 작업

- 제품군 선정 (디지털 카메라, TV)
- 제품군의 기능들 식별 (identifying features)
  - Product feature sets : 두 분야에서 특징 set을 선별
  - Synonym sets: 도메인 지식에 따라 동의어 set을 만들어서 처리
- Test-set : 특정 웹사이트의 디지털 카메라, TV 리뷰에서 테스트 용으로 1,000개의 리뷰 문장을 무작위로 가져옴
- Data-set : 1,516,001개의 문장 추출 (그 중 특징이 명확이 명시된, 약 16%만을 사용)

# 자연어 처리

- 리뷰를 문장화 / 토큰화
- 품사 태깅, 불용어 처리

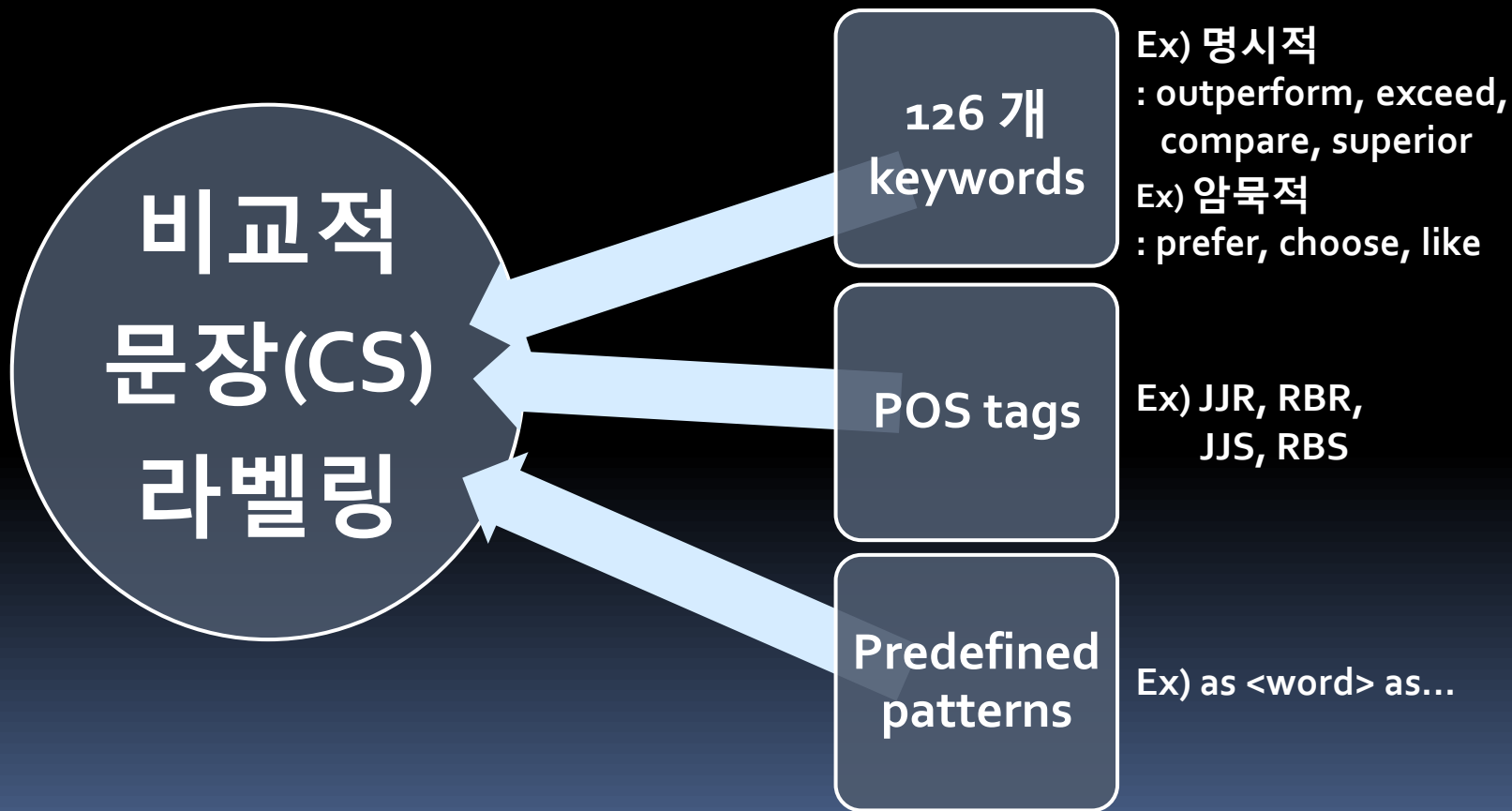
# 라벨링

- 주관적 문장 (SS) 라벨링
  - : 직접적인 칭찬 또는 비평을 나타내는 문장
  - Positive Subjective(PS)
    - (ex) This camera has **great** picture quality and **conveniently** priced.
  - Negative Subjective(NS)
    - (ex) The picture quality of this camera is **really bad**.

→ 방법론적인 의논이 필요!!

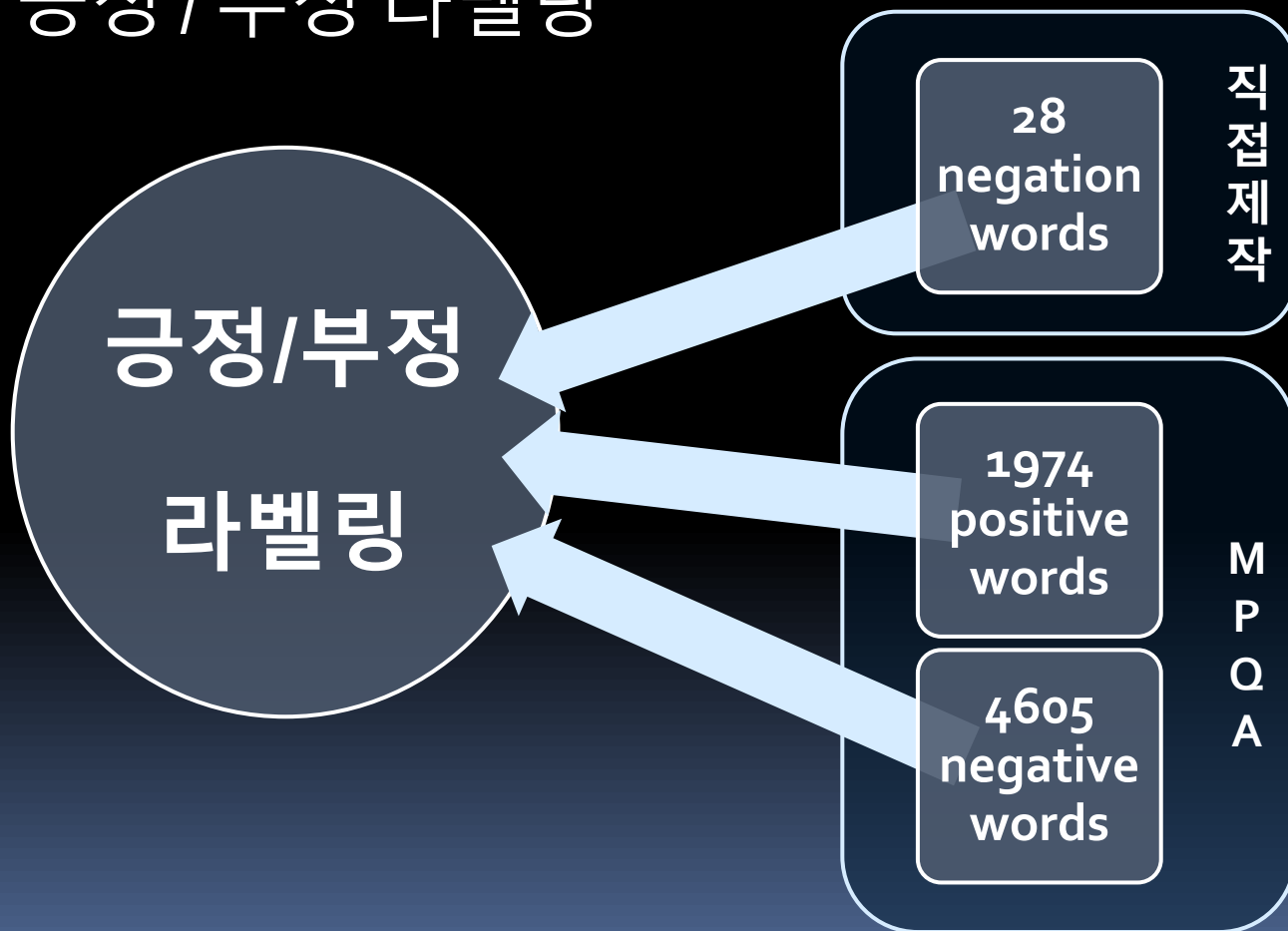
# 라벨링

- 비교적 문장 (CS) 라벨링



# 라벨링

- 긍정 / 부정 라벨링





# 그래프 생성

- TextRank의 방법론을 차용

## 용어의 정의

---

**\*노드 (Vertex)**

: 제품을 의미

**\*노드 간 엣지**

: 제품 간 비교를 언급하는 문장을 기반으로 계산

# pRank 적용

- 각 노드들의 스코어와 노드 간 엣지의 가중치를 구하고, 반복적인 연산 수행 후 각 노드들의 스코어, 즉 각 제품들의 스코어를 통해 상대적 품질을 알아내는 것으로 논문은 끝남

# 논문 구현 작업 FLOW

