**Knowledge graph refinement: A survey of approaches and evaluation methods**

**Abstract**

DBpedia, YAGO, Freebase 등의 open KG가 있다. KG는 Semi-structured knowledge로 구성(Wikipedia) / 통계적, 언어적 방법의 결합으로 웹에서 얻어짐.

최종적 산출물 = KG(completeness와 correctness 사이의 trade-off를 잘 맞추면서)

KG의 Utility를 높이기 위해서 refinement methods!(infer and add missing knowledge to graph/ 정보의 잘못된 부분을 밝혀냄) => KG Refinement approach!

**1. Introduction**

KG는 Domain-specific or domain-independent 할 수 있다. 그래프로써 세상의 일반적인 지식이 representation 되는 중.

Cyc => curated! / Freebase, Wikidata => 대중 참여로 edit! / Wikipedia 같은 large-scale, semi-structured knowledge base!로부터 추출된 DBpedia and YAGO.

또한 추출(extratction) 방법도 제안되고 있다.(unstructured or semi-structured information 으로 부터) => NELL, PROSPERA, KnowledgeVault 가 추출 방법들

하지만 아직 전부다 not perfect! 세상의 모든 entity의 정보를 같긴 힘들다(full coverage 불가) = 많은 정보가 graph로 확장! = 그래프가 커짐 // 또한 KG가 fully correct 해질 수 없다.(그저 heuristic method를 쓸 때) => coverage와 correct는 trade off! 이것을 잘 조절하는 것이 중요

많은 논문들은 기존의 KG를 이용하는 것이 아닌 자신들이 새롭게 KG construction! 여기서는 construction 아니고 KG refinement (기존 것을 향상시킴)

기존 것들은 이미 어느정도 기술이 있음(adding missing knowledge, identifying and removing errors) = 대개 KG의 정보를 directly 이용(automatic approach를 위한 training info로 사용?) 이부분 이해 안감.

따라서 뭔가 비슷하지만 다르다. (refinement는 given graph가 필요)

많은 refinement step이 graph 생성할 때 요구된다. Logical reasoning => validating consistency (statements의 <술부>), 그리고 inconsistent statements는 제거 = 이런게 post processing = 초기 KG construction 후의 적용! = refinement methods

Construction과 refinement는 다른 이점 => RF는 임의의 KG에서 작동하며 향상시킴/ RF는 heuristic fine-tune 보다 더 큰 효과

Evaluating RF는 CR과 독립적으로 effect의 효과적 이해 => 제안된 approach를 좀 더 향상!?

Sec 2 – KG의 Semantic Web(SW)에서의 소개

Sec 3, 4 – approach, method

Sec 5, 6 – completion method review (increasing coverage) / error detection (increasing correctness) in KG

….

**2. Knowledge graphs in the Semantic Web**

Semantic Web => RDF 표준<웹상의 자원의 정보를 표현하기 위한 규격>을 이용해 graph-based representation!

Entity => node / relation => edge

Entity는 type을 가짐(relation에 의해 나타내어짐, 관련이 있다는 소리)

가능한 type과 relation의 sets은 schema or ontology(type과 relation 사용의 상호 관계, 규제를 정의)에 의해 구성됨.

Linked Data => SW의 다른 dataset들 연결 => interlinking으로 1000 datasets interlinked (two dataset에서 identical entities connect!)??

KG => 1. 현실 세계 정보 반영 / 2. 스키마에서 가능한 class 정의 / 3. 임의의 엔티티를 서로 interrelating? / 4. 다양한 domain 다룸

애매한 단어들

시맨틱 웹, RDF dataset, RDF standard, 온톨로지, knowledge base, DBpedia, YAGO

A-Box, T-Box, Linked Data