

대용량 그래프에서의 유사 매칭을 위한 그래픽 사용자 인터페이스 기반 서브 그래프 생성 도구에 대한 연구

송재오^{O*}, 홍승민^{**}, 이상문^{***}

^{O*}㈜케이아이씨티 기업부설연구소

^{**}서원대학교 컴퓨터공학과

^{***}한국교통대학교 컴퓨터정보공학과

e-mail: jos@k-ict.co.kr^{O*}, globalsh1@gmail.com^{**}, smlee@ut.ac.kr^{***}

A Study on GUI based Subgraph Generation Tool for Similar Matching in Large Capacity Graphs

Je-O Song^{O*}, Seung-Min Hong^{**}, Sang-Moon Lee^{***}

^{O*}R&D Institute, K-ICT Co.,Ltd.

^{**}Department of Computer Science and Engineering, Seowon University

^{***}Dept. of Computer Sci. & Info. Eng., Korea Nat'l Univ. of Transportation

● 요약 ●

최근 빅데이터를 비롯한 각종 실험 장비의 발전에 따라 첨단 분야에서의 과학데이터가 급격히 증가하고 있는 가운데, 그래프 매칭은 컴퓨터 네트워크 모니터링, 소셜 네트워크의 진화 분석, 생물학 네트워크에서 모티프(motif) 탐지 등 네트워크 분석 및 데이터 마이닝 분야에서 널리 활용되고 있다. 이와 같이, 폭발적으로 증가하는 데이터에 대한 네트워크 모델링 및 유사 그래프 매칭 분석을 수행하기 위한 연구 및 기반 기술 개발은 필수적인 실정이다. 본 논문에서는 이미 확보된 대용량 그래프에서 유사한 형태의 서브 그래프를 매칭할 수 있는 GUI(Graphic User Interface)기반의 생성 도구를 제안한다.

키워드: 서브(Sub), 대용량(Large Capacity), 그래프(Graph), 유사(Similar), 매칭(Matching)

I. Introduction

스마트 기기의 보급과 발달로 스마트 시대를 지나 초연결 지능 연결 형태의 4차 산업 혁명 시대로 진입하고 있는 가운데 이미 여러 분야에서 빅데이터를 이용한 인공지능 서비스 등의 데이터 기반 기술에 대한 연구들이 활발히 진행되고 있다. 유전자 정보에 대한 방대한 데이터에 대해서도 대용량 그래프 작성 등을 이용한 방법들이 사용되고 있는 등 각종 실험 장비의 발전에 따라 첨단 분야에서의 과학데이터가 급격히 증가하고 있으며, 그래프 형태로 생산되는 데이터가 급격히 증가하는 추세이다. 이러한 환경에서 대상이 되는 대용량 그래프로부터 질의로 주어진 서브 그래프를 검색하여 찾아주는 서브 그래프 매칭이 주목을 받고 있다. 완벽하게 일치하는 결과만을 검색하는 완전 일치 서브 그래프 매칭에 비해 유사 서브 그래프 매칭은 질의 그래프와 비슷한 결과까지 검색하기 때문에 다양한 분야의 응용 및 연구에서 활용된다. 본 논문에서는 대용량 그래프에서 Command 형태의 Text 기반 서브 그래프 생성하는 기존의 방식을 Graphic 형태로 생성하고 시각적으로 확인하여 대용량 그래프 데이터에서 유사 매칭 그래프 데이터를 찾아낼 수 있는 방법을 제안한다.

II. Design and Implementation

본 논문에서 제안하는 방법은 서브 그래프를 GUI 형태로 생성하는 것을 시작으로 그림1과 같이, JSP를 사용하여 웹에 기반한 그래프 작성을 할 수 있도록 하였으며 Drag and Drop 방식으로 동작한다.

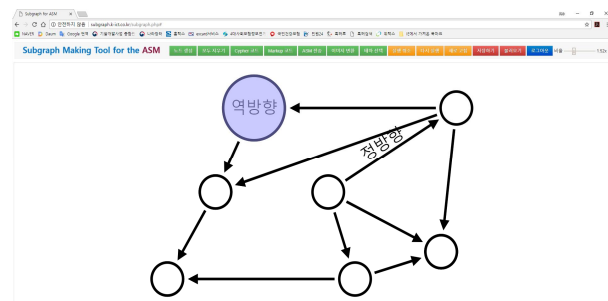


Fig. 1. Drag&Drop을 이용한 노드 생성 및 연결

해당 GUI 기반의 서브 그래프 생성 도구는 다중 노드를 작성하고 각 노드의 연결과 수정을 할 수 있도록 하였다. 즉, 단순히 시각적인 모양만을 작성하는 것이 아니라 노드 네임과

속성 뿐만이 아니라 Vertex, Edge의 관계 타입과 그에 대한 속성 값을 설정할 수 있다. 이는 그림2와 같이 Cypher 코드로서의 표현도 가능하게 해준다.

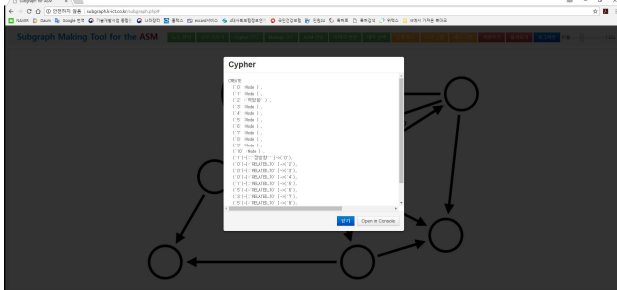


Fig. 2. 생성된 서브 그래프에 대한 Cypher 코드

또한, 이와 같은 코드 변환으로 그림3과 같이 Neo4j와 연계되어 현재 그래프에 대한 구동 상태를 확인하고 데이터에 대한 질의 처리도 실행된다.

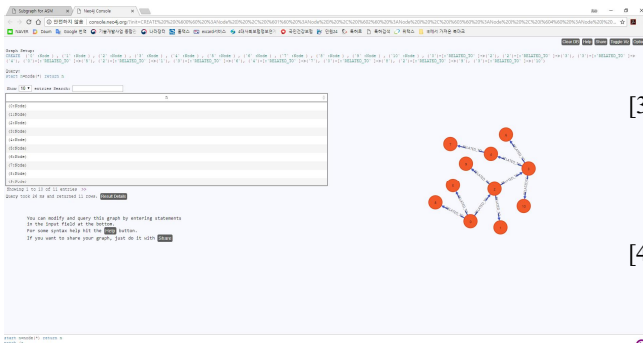


Fig. 3. Neo4j 콘솔 연동

상기와 같은 일련의 과정을 통한 서브 그래프 데이터는 그림4와 같이 대용량 그래프 데이터에서 매칭되어 특정 패턴에 대한 분석을 가능하게 해준다.

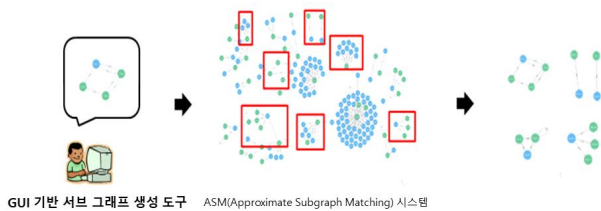


Fig. 4. 대용량 그래프 데이터에서의 서브 그래프 매칭

본 논문에서는 기존의 방식을 Graphic 형태로 생성하고 시각적으로 확인이 가능한 서브 그래프를 통해 대용량 그래프 데이터에서 정확한 질의를 통한 정확한 유사 매칭 그래프 데이터에 결과를 확인할 수 있게 해준다.

REFERENCES

- [1] Jongtae Lim, Minyoung Kim, Dojin Choi, Dongmin Seo, Seok Jong Yu, Kyoungsoo Bok, Jaesoo Yoo, "Approximate Sub-Graph Matching Scheme for Large-Scale Graph Data in Spark Environments", Proceedings of KIISE Conference, Vol.12, pp.200-202, 2017.
- [2] Yeo-Been Yoon, Jung-Hyun Cho, Je-O Song, Sang-Moon Lee, "Design and Implementation of Manufacturing Resource Management System as Open Type based on Mold Image Data", Proceedings of KSCI Conference, Vol.25, No.1, pp.227-228, 2017.
- [3] Jeo Song, Jin-Hwan Jeon, Un-Kyung Song, Sang-Moon Lee, "Design and Implementation of a Mobile Search Method based on Images", Proceedings of KSCI Conference, Vol.24, No.1, pp.33-36, 2015.
- [4] Jeo Song, Yong Goo Park, Sang Moon Lee, "A Study on the Related Information of Smart Design Contents for Service Design", Proceedings of the MITA International Conference, Vol.11, pp.241-243, 2015.
- [5] Talukder, Nilothpal, and Mohammed J. Zaki. "A distributed approach for graph mining in massive networks." Data Mining and Knowledge Discovery, Vol.30, No.5, pp.1024-1052, 2016.

III. Conclusions

대용량 그래프에서 서브 그래프에 대한 일치성을 판단하기 위해 대부분의 기존 방법들은 각 노드에 대한 속성들을 Command 형태로 작성하여 사용하여 왔다. 이는 작성중인 서브 그래프에 대한 시각적 확인이 어렵기 때문에 질의어로서의 정확한 입력 확인이 어려웠다.