

시각화 기반 모바일 라이프 로그 시맨틱 네트워크 연관 검색 시스템

(An Associative Search System for Mobile Life-log Semantic Networks based on Visualization)

오근현[†] 김용준^{**}
(Keunhyun Oh) (Yongjun Kim)

조성배^{***}
(Sung-Bae Cho)

요약 최근 모바일 기기의 다양한 센서들을 통하여 데이터를 수집하여 개인의 삶을 기록하는 연구들이 진행 중이다. 효율적인 모바일 라이프로그의 저장과 탐색을 위해 연관 검색이 가능한 모바일 라이프로그 시맨틱 네트워크가 제안되었다. 기존의 시맨틱 네트워크 상의 검색은 텍스트 기반으로 관계를 바탕으로 하는 사용자의 능동적인 연관 검색에 한계가 있었다. 본 논문에서는 모바일 라이프로그의 연관 검색을 위해 검색 과정과 결과를 시각화된 네트워크로 데이터간의 관계를 보여주는 선택 연관 검색과 키워드 연관 검색을 제안한다. 복잡한 시맨틱 네트워크에 대해서 사용자의 이해도를 높이기 위해 의미 추상화를 적용하였다. 실제 모바일 라이프로그 시맨틱 네트워크 데이터를 바탕으로 질의를 해결하는 과정을 기존 연구에서 보였던 텍스트

기반 연관 검색 방법과 비교하고 사용성 평가를 시행함으로써 유용성을 입증하였다.

키워드 : 모바일 로그, 시맨틱 네트워크, 연관검색, 네트워크 시각화, 의미추상화

Abstract Recently, mobile life-log data are collected by mobile devices and used to recode one's life. In order to help a user search data, a mobile life-log semantic network is introduced for storing logs and retrieving associative information. However, associative search systems on common semantic networks in previous studies provide for a user with only found data as text to users. This paper proposes an associative search system for mobile life-log semantic network that supports selection and keyword associative search of which a process and result are a visualized graph representing associative data and their relationships when a user inputs a keyword for search. In addition, by using semantic abstraction, the system improves user's understanding of search result and simplifies the resulting graph. The system's usability was tested by an experiment comparing the system and a text-based search system.

Key words : Mobile Log, Semantic Networks, Associative Search, Network Visualization, Semantic Abstraction

1. 서론

최근 모바일 기기가 발달함에 따라 다양한 센서가 내장되어 개인의 일상에 대한 정보인 모바일 라이프로그의 수집이 가능하게 되면서 이를 바탕으로 개인의 삶을 기록하는 연구들이 진행 중이다. 모바일 라이프로그의 효과적인 관리와 검색을 위하여 데이터간의 관계를 바탕으로 검색을 수행할 수 있는 모바일 라이프로그 시맨틱 네트워크가 제안하였다[1].

사람은 기억을 회상할 때 정보 사이의 관계성을 바탕으로 연관된 사실을 떠올린다[2]. 기존의 시맨틱 네트워크 상에서의 연관검색은 검색 결과로 데이터만을 제공하고 관계를 적절히 제공하지 않아 사용자가 직접 데이터를 검색할 시에는 검색 결과에 대한 이해도, 만족도, 검색의 효율성이 떨어진다. 따라서 데이터간의 관계를 직관적으로 알 수 있도록 시각화된 네트워크 상에서 검색을 수행하고 연관 검색 결과를 시각화된 그래프로 제공하는 것이 유용하다.

본 논문에서는 모바일 라이프로그의 검색을 위한 시각화 기반의 모바일 라이프로그 시맨틱 네트워크 연관 검색 시스템을 제안한다. 시각화된 그래프 기반의 연관 검색 방법과 검색 결과 그래프의 분석을 용이하게 하기 위해 의미 추상화(Semantic abstraction)를 시맨틱 네트워크에 적용하였다.

· 이 논문은 2009년도 정부(교육과학기술부)의 재원으로 한국연구재단-신기술융합형 성장동력사업의 지원을 받아 수행된 연구임(No. 2009-0093676)
· 이 논문은 제36회 추계학술발표회에서 '의미 추상화와 키워드 검색 기반의 시맨틱 네트워크 연관 검색 시스템'의 제목으로 발표된 논문을 확장한 것임

† 학생회원 : 연세대학교 컴퓨터과학과
ocworld@sclab.yonsei.ac.kr
** 정 회 원 : 연세대학교 컴퓨터과학과
yjkim@sclab.yonsei.ac.kr
*** 종신회원 : 연세대학교 컴퓨터과학과 교수
sbcho@cs.yonsei.ac.kr
논문접수 : 2009년 12월 23일
심사완료 : 2010년 3월 4일

Copyright©2010 한국정보과학회 : 개인 목적이나 교육 목적인 경우, 이 저작물의 전체 또는 일부에 대한 복사본 혹은 디지털 사본의 제작을 허가합니다. 이 때, 사본은 상업적 수단으로 사용할 수 없으며 첫 페이지에 본 문구와 출처를 반드시 명시해야 합니다. 이 외의 목적으로 복제, 배포, 출판, 전송 등 모든 유형의 사용행위를 하는 경우에 대하여는 사전에 허가를 얻고 비용을 지불해야 합니다.

정보과학회논문지: 컴퓨팅의 실제 및 데이터 제16권 제6호(2010.6)

2. 관련연구

2.1 시맨틱 네트워크 검색

Liu 등은 일반 상식을 시맨틱 네트워크로 구성하고 지식 표현과 연관 검색이 가능한 어플리케이션들을 소개하였다[3]. Rinaldi는 웹 상에서 웹 문서상의 단어 사이의 관계를 바탕으로 한 시맨틱 네트워크를 구성하고 검색을 수행하는 방법을 제안하였다[4]. 이러한 연구에서의 검색은 데이터 간의 관계를 제공해주지 않아서 사용자가 검색 결과와 질의어가 어떤 관계를 가지는지 이해하기 어렵다.

2.2 네트워크 시각화

Jia 등은 전체 네트워크상의 각 노드의 중요성과 에지의 의미를 변화시키지 않고 그래프를 간략화하기 위하여 중요성이 떨어지는 에지를 제거하고 정보를 클러스터링 하였다[5]. 이처럼 대부분의 네트워크 시각화 연구는 분석이 목적이므로 데이터의 유실을 가져온다. 검색을 목적으로 할 경우에는 실제 필요한 데이터가 유실될 가능성이 있다. Shen 등은 소셜 네트워크를 분석하기 위해 의미 추상화를 제안하였다. 의미 추상화는 전체 네트워크의 구조를 파악할 수 있는 온톨로지 그래프(ontology graph)와 불필요한 타입의 노드를 보여 주지 않는 유도그래프(induced graph)의 구성을 의미한다[6]. 본 논문에서는 의미 추상화를 시맨틱 네트워크에 적합하게 구성하여 적용하였다.

3. 제안하는 방법

3.1 모바일 라이프로그 시맨틱 네트워크

모바일 라이프로그 시맨틱 네트워크는 모바일 라이프로그의 효과적인 저장과 검색을 위해 이전연구에서 제안되었다. 활동이론(activity theory)에 따르면 행동이 모든 상황의 단위가 되어야 한다. 따라서 행동에 영향을 주는 요소인 장소와 시간, 그 행동을 할 때 발생이 이벤트인 모바일 기기의 기능을 행동 중심으로 구성하였다. 표 1은 각 노드의 종류에 대한 정의를 보여준다[1].

3.2 연관 검색 시스템 개요

제안하는 검색 시스템의 개요는 그림 1과 같다. 먼저 모바일 라이프로그 시맨틱 네트워크를 시각화한다. 시각화된 시맨틱 네트워크 상에서 선택 연관검색과 키워드 연관검색을 통해 검색을 수행할 수 있다. 시스템은 결과를 시각화된 결과그래프로 제공한다. 결과그래프가 복잡하면 이해가 어렵기 때문에 의미추상화를 사용한다. 의미 추상화는 시맨틱 네트워크의 구조에 대한 정보를 온톨로지 그래프를 통하여 사용자에게 제공한다. 그리고 온톨로지 그래프 상에서 선택된 타입의 노드들과 그 노드들 사이의 관계들로부터 구성된 유도그래프를 통해 연관검색 결과로 제공되는 그래프의 간략화가 가능하다.

표 1 모바일 라이프로그 시맨틱 네트워크 노드 종류

노드타입	정의
활동	근노드
행동카테고리	행동의 분류
행동	각 카테고리로 나뉜 행동
장소	GPS기반으로 레이블링한 데이터
시간	모바일 기기 이벤트 발생 시간
모바일기기 기능	모바일 기기의 기능
내용	문자내용, 음악파일명, 사진이미지
상대방	전화/문자를 주고 받은 상대방

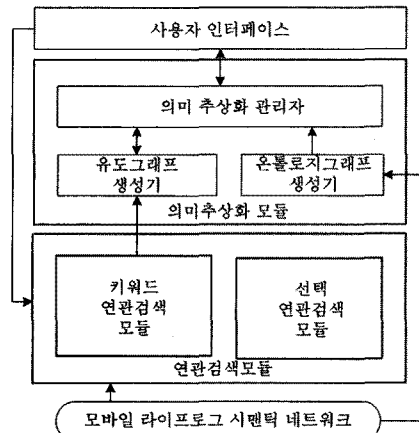


그림 1 연관검색 시스템 개요

3.3 연관검색

본 논문에서는 효율적인 연관 검색을 위해 시각화된 그래프를 사용하여 사용자에게 데이터 간의 관계를 제공한다. 이전연구의 선택 연관검색을 적용하였다. 데이터를 선택하면 에지로 연결된 노드들을 보여준다. 검색을 수행할 시에는 부분적인 정보만을 가지고 있더라도 검색을 수행하면서 연관된 정보를 차례로 추가 정보를 얻고 다음 단계의 검색을 수행할 수 있다[1]. 키워드 연관검색을 제안한다. 정확히 알고 있는 키워드가 있을 경우 그 키워드를 질의어로 입력하면 시스템은 노드와 연관된 조상/자손 노드를 포함한 검색 결과그래프를 출력으로 한다. 시스템은 질의어를 깊이 우선탐색 알고리즘을 사용하여 찾는다. 입력한 키워드를 찾게 되면 해당 노드와 근노드에서 해당 노드까지의 경로상의 노드와 관계를 결과그래프에 추가한다. 또 키워드 노드의 모든 자손 노드를 결과그래프에 추가한다. 키워드 연관검색의 의사코드는 그림 2와 같다. 시각화된 네트워크 상에서 검색을 수행함으로써 검색 결과에 대한 이해와 만족도를 향상시킨다.

3.4 시맨틱 네트워크 의미 추상화

결과그래프의 규모가 커지고 복잡성이 높아지면 사용

```
function Graph KeywordAssociativeSearch
(string keyword, Graph S)
  Graph ResultGraph;
  Node KeywordNode
  = DFS(keyword,S).Result();
  ResultGraph.add(
  DFS(keyword,S).Routes());
  ResultGraph.add(
  DFS(Keyword,S).Result());
  ResultGraph.add(
  Traverse(keywordNode,S).Routes());
  return ResultGraph;
end
```

그림 2 키워드 연관검색 의사 코드

자가 이해하기 어렵다. 결과그래프의 의미를 사용자에게 효과적으로 전달하기 위하여 Shen 등이 제안한 의미 추상화를 사용하였다. 전체 네트워크의 구조와 연관성을 보여주는 온톨로지 그래프를 통하여 원하는 종류의 노드들과 관계들을 보여주는 유도그래프를 생성하는 과정을 의미한다[6]. 온톨로지 그래프는 $OG=(T_v, T_E)$ 구성된다. T_v 는 노드에 해당하며 시맨틱 네트워크 상의 노드의 타입을 의미한다. T_E 는 에지에 해당하고 시맨틱 네트워크 상의 에지의 타입을 의미한다[3]. T_v 를 타입노드라고 명명하고 T_E 를 타입에지라고 명명하겠다. 본 논문에서는 온톨로지 그래프의 생성규칙을 시맨틱 네트워크에 적합하게 새롭게 정의하였다. 그림 3의 우측 부분은 모바일 라이프로그 시맨틱 네트워크의 온톨로지 그래프의 예시를 보여준다.

타입노드의 크기는 식 (1)을 따른다. N_i 는 i 번째 노드를 의미하며 R_i 는 i 타입노드의 반지름을 의미한다. 상수 C 는 사용자가 지정한 최대 타입노드 크기이다. NC_i 는 시맨틱 그래프 상에서 타입노드와 매핑되는 노드의 숫자를 의미한다. EC_{max} 는 시맨틱 네트워크 상에서 가장 많은 타입의 노드를 의미한다. 시맨틱 네트워크 상에서 그 타입에 해당하는 노드가 얼마나 많이 있는지에 따라서 결정된다.

$$R_i = C * \frac{NC_i}{NC_{max}} \quad (1)$$

타입에지의 두께는 식 (2)를 따른다. E_i 는 i 번째 에지를 의미하고 E_i 는 i 에지의 두께를 의미한다. E_i 는 사용자가 지정한 최대 타입에지의 두께이다. EC_i 는 시맨틱 네트워크 상에서 출발 노드 타입과 도착 노드 타입 사이의 에지의 수를 의미한다. EC_{max} 는 타입과 타입 사이 에지의 수 중 최대 에지의 수를 의미한다.

$$E_i = C * \frac{EC_i}{EC_{max}} \quad (2)$$

타입에지의 색은 RGB를 바탕으로 식 (3)을 따랐다.

$$\begin{cases} Red = 255 - 255 \left(\frac{EC_i}{EC_{max}} \right) \\ Green = \frac{Red}{2} \\ Blue = \frac{Green}{2} \end{cases} \quad (3)$$

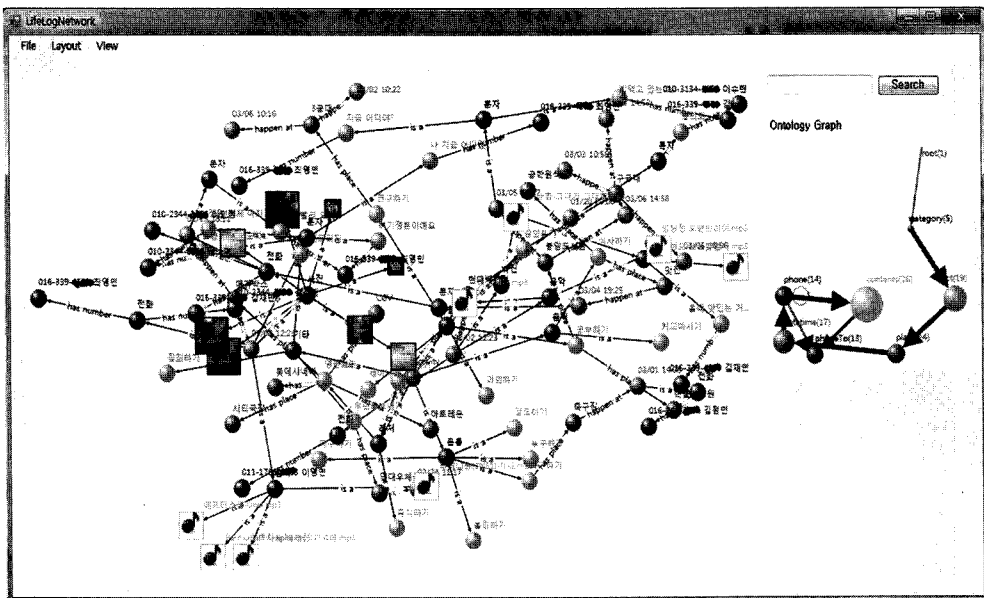


그림 3 구현된 모바일 라이프로그 시맨틱 네트워크 연관 검색 시스템. 왼쪽은 사용자가 선택 연관 검색을 행할 수 있고 검색 결과 그래프를 확인할 수 있다. 우측 상단은 키워드 연관 검색에서 질의어를 입력하는 부분이다. 또한 우측에는 온톨로지 그래프가 있으며 노드를 선택하면서 유도 그래프를 생성할 수 있다.

더 많은 에지를 포함하고 있는 타입에지는 어두운 색으로 표시하여 온톨로지 그래프를 보았을 때 직관적으로 구조와 각 타입 사이에 얼마나 많은 관계들이 있는지를 판단할 수 있도록 하였다.

온톨로지 그래프에서 타입노드를 선택하면 시맨틱 네트워크에서 해당 타입의 노드들이 삭제되며 삭제된 타입노드를 다시 선택하면 그 타입의 노드들은 다시 나타난다. 복잡한 그래프에서 검색에 불필요한 타입의 노드들을 제거하여 그래프를 간략화시켜 유도 그래프를 생성한다. 온톨로지 그래프와 유도 그래프는 사용자의 결과그래프에 대한 분석과 이해를 돕는다.

4. 실험 및 평가

제안하는 시스템은 C#으로 구현되었으며 그래프 시각화에는 마이크로소프트의 NodeXL 라이브러리를 사용하였다(<http://www.codeplex.com/NodeXL>). 그림 3은 구현된 시스템의 전체 모습을 보여준다. 실험에서 사용된 모바일 라이프로그 시맨틱 네트워크는 스마트폰 삼성 m-4650와 GPS BT-335를 사용하여 3일 동안 수집한 대학생의 모바일 로그 데이터를 바탕으로 만들어졌으며 109개의 노드와 106개의 에지로 구성되었다.

4.1 연관검색

주어진 질의어는 “영화를 보고 있었을 때 중요한 문자를 받았다. 어떤 영화인지 언제, 누구에게 받은 문자인지는 알 수 없다. 어떤 문자인가?”이다. 사용자는 온톨로지 그래프를 통해 시맨틱 네트워크의 구조를 파악하고 근노드 활동을 선택하고 검색을 시작한다(그림 4(a)). 행동카테고리가 연상되고 ‘레지’를 선택하고 영화 보기를 선택한다. 관련 장소들이 상기되는데 M영화관에서 기록된 시간을 보고 문자 내용을 확인한다(그림 4(b)). 그림 5는 키워드 연관검색을 통해 질의어 ‘문자’와 연관된 노드를 결과그래프 상에서 확인할 수 있다.

4.2 의미추상화

온톨로지 그래프는 그림 2에서 볼 수 있다. 시맨틱 네트워크의 구조와 어떤 타입의 노드가 더 많이 존재하는지 어떤 타입들끼리 더 많이 연결되어 있는지를 판단할 수 있다. 그림 6은 온톨로지 그래프를 통하여 ‘상대방’ 타입과 ‘내용’ 타입을 제거한 유도그래프이다. 복잡한 그래프가 간결화되어 원하는 데이터를 찾기에 용이하다.

4.3 평가

모바일 라이프로그 시맨틱 네트워크 상에서 주어진 5개의 질의를 사용자에게 검색하도록 하였다. 검색의 효율성을 측정하기 위해 기존 연구와 같이 연관 검색 결과로 텍스트형태로 제공하는 연관검색 방법과 제안하는 방법의 탐색과정 수와 연관되는 데이터 수를 비교하였다. 평균 데이터 획득 수는 사용자가 하나의 상황으로 이

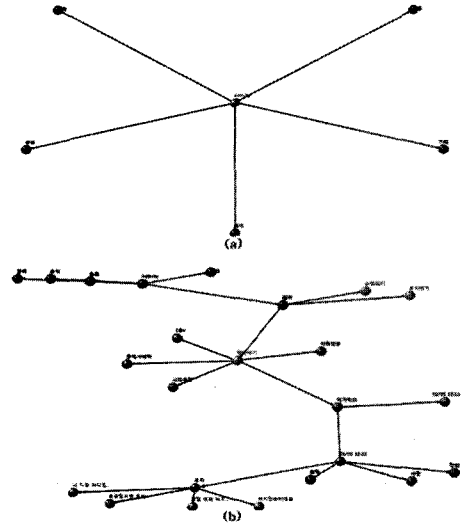


그림 4 선택 연관검색 예시

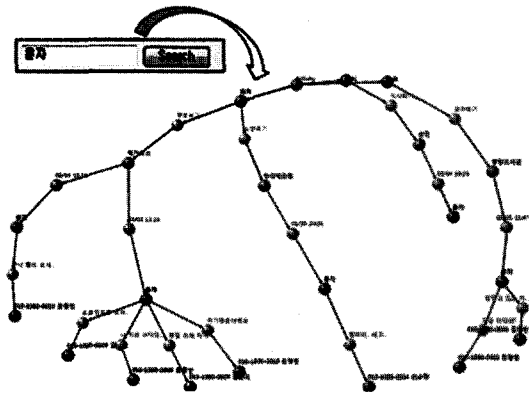


그림 5 키워드 연관검색 예시(키워드: 문자)

해한 데이터의 수이다. 표 2에서 기존 방식에 비해 제안하는 방법의 평균 탐색과정 수가 적은 것을 보여준다. 관계까지 제공하기에 대부분 1~2번의 검색으로 정보를 찾을 수 있었다. 검색 결과에 대한 이해에 도움을 준다면 평균 연관 데이터 획득 수가 많은 것은 유용한 부분이다. 사용자의 연관 데이터 이해도의 측정을 위해 표 3의 문항으로 사용성 평가를 시행하였다. 7명의 사용자를 대상으로 기존 방법과 제안하는 시스템을 비교하였다. 이 문항들은 연관 검색 결과의 이해도, 만족도, 그리고 유용성을 평가한다. 각 사용자별 비교 평가 결과의 평균은 그림 7과 같다. 또한 표 4의 SUS문항을 사용하여 사용자의 시스템에 대한 만족도를 조사하였다. 점수는 1~5점으로 1점은 ‘매우 아니다’, 5점은 ‘매우 그렇다’이다. 그림 8은 SUS평가의 결과를 보여준다. 제안하는 시스템이 사용자 만족도가 높은 것을 보여준다.

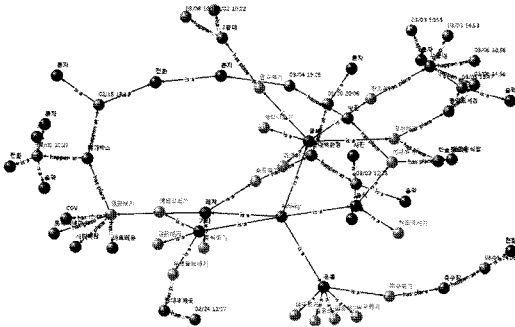


그림 6 의미 추상화를 통해 유도된 그래프 ('상대방'과 '내용'타입 제거)

표 2 평균 탐색과정 수와 평균 연관 데이터 획득 수 비교

	텍스트 기반	제안하는 방법
탐색과정 수	2.2	1.2
평균데이터획득수	8.2	13.2

표 3 검색 시스템 비교 사용성 평가 문항

번호	문항
1	제공하는 검색기능은 유용하였습니까?
2	화면에 보여주는 검색결과가 유용하였습니까?
3	검색 질문의 검색결과에 만족하셨습니까?

표 4 SUS 평가 문항

번호	문항
1	나는 이 시스템을 자주 사용할 것 같다.
2	시스템에 불필요하게 복잡한 부분이 있다.
3	시스템이 사용하기 쉽다고 생각한다.
4	이 시스템을 사용하기 위해서 전문가가 필요할 것 같다.
5	이 시스템은 다양한 기능이 조직적으로 잘 결합되어 있다.
6	이 시스템은 너무 불안정한 것 같다.
7	대부분의 사람이 이 시스템의 사용방법을 빨리 익힐 것이다.
8	이 시스템은 사용하기 귀찮은 부분이 있다.
9	나는 이 시스템을 사용했다는데 자부심이 생긴다.
10	이 시스템을 계속 사용하기 위해서는 많은 것을 배워야 할 것 같다.

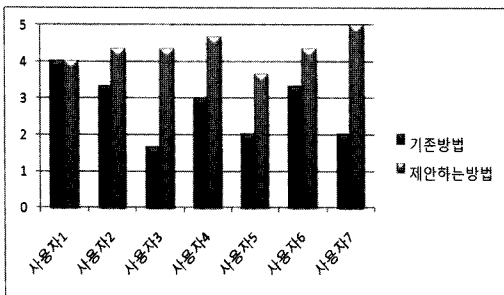


그림 7 검색 시스템 유용성 비교 사용성평가 결과(평균)

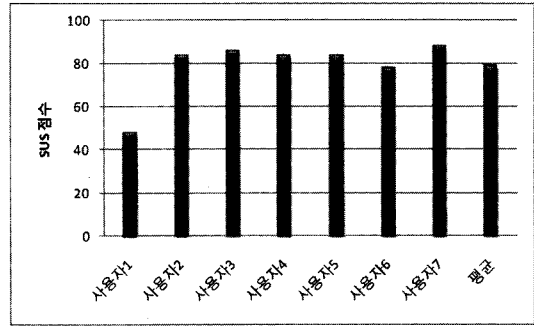


그림 8 SUS 점수

5. 결론 및 향후 연구

본 논문에서는 모바일 라이프로그를 효과적으로 검색하기 위한 시각화 기반의 모바일 라이프로그 시맨틱 네트워크 연관검색 시스템을 제안하였다. 검색 과정과 결과에서 데이터간의 관계를 제공하기 위해 네트워크를 시각화한 선택 연관 검색과 키워드 연관 검색을 제안하여 사람의 기억 회상과 유사한 검색이 가능하게 하였다. 시각화한 네트워크의 복잡성을 줄이고 사용자의 이해를 돕기 위해 의미 추상화를 적용하였다. 시스템을 구현하고 실제 데이터를 기반으로 텍스트 기반 검색 방식과 질의 해결 능력을 비교하고 사용성 평가를 시행하여 시스템이 검색 효율성과 사용자의 이해와 만족을 높임을 입증하였다.

모바일 라이프로그 시맨틱 네트워크를 이용하여 개인의 삶을 분석하는 연구가 향후 연구로 진행할 예정이다.

참고 문헌

- [1] K. Oh, Y. Kim, and S.-B. Cho, "A mobile life-log search interface by associative traverse of semantic network," *Proc. of KCC*, vol.36, no.1(B), pp. 193-197, 2009.
- [2] J. G. Raaijmakers, and R. M. Shiffrin, "Search of associative memory," *Psychological Review*, vol.88, no.2, pp.93-134, 1981.
- [3] H. Liu, and P. Singh, "ConceptNet-a practical commonsense reasoning tool-kit," *BT Technology Journal*, vol.22, no.4, pp.211-226, 2004.
- [4] A. M. Rinaldi, "An ontology-driven approach for semantic information retrieval on the web," *ACM Trans. on Internet Technology*, vol.9, no.3, 2009.
- [5] Y. Jia et al., "On the visualization of social and other scale-free networks," *IEEE Trans. on Visualization and Computer Graphics*, vol.14, no.6, pp.1285-1292, 2008.
- [6] Z. Shen et al., "Visual analysis of large heterogeneous social networks by semantic and structural abstraction," *IEEE Trans. on Visualization and Computer Graphics*, vol.12, no.6, pp.1427-1439, 2006.