

컨셉넷과 키그래프를 이용한 일상생활 요약

이영설⁰ 김경중 조성배
 연세대학교 컴퓨터과학과
tiras@sclab.yonsei.ac.kr, {kjkim, sbcho}@cs.yonsei.ac.kr

Summarizing User's Daily Life with ConceptNet and KeyGraph

Youngseol Lee⁰, Kyung-Joong Kim, Sung-Bae Cho
 Department of Computer Science, Yonsei University

요 약

모바일 기기에서 수집된 데이터를 바탕으로 사용자의 기억을 되살리거나 사용자 맞춤형 서비스를 제공하기 위해서는 사용자가 하루 동안 수행한 수많은 행동이나 겪은 사건들을 기록할 필요가 있다. 그러나 사용자의 하루 동안 발생한 모든 일을 보여주기 보다는 사용자에게 있어서 중요한 일만 보여주는 것이 사용자의 행동을 분석하고 사용자에게 맞는 서비스를 제공하는 데 더 도움이 될 것이다. 이전의 연구에서는 키그래프를 이용하여 사용자의 하루를 요약하려고 시도하였으나 사용자로부터 얻은 데이터에 직접 키그래프를 적용한 결과로 얻은 데이터는 사람이 직관적으로 그 중요성을 이해하기 힘들었다. 이를 해결하기 위해 보다 상위 수준의 정보에 키그래프를 적용하였다. 본 논문에서는 스마트 폰에서 수집된 GPS 위치 정보를 장소정보로 변경하여 컨셉넷에 입력하고 관련된 개념을 추출한다. 컨셉넷(ConceptNet)은 수많은 어휘들의 연관관계를 의미망 형태로 표현한 것으로 사람이 일상생활에서 이용하는 상식을 포함하고 있다. 컨셉넷에서 추출된 개념들로 문서를 생성하고 생성된 문서에 키그래프 알고리즘을 적용하여 사용자가 수행한 행동에 관련된 개념 중 핵심적인 개념을 추출한다. 이렇게 추출된 개념들은 사용자가 하루 동안 수행한 행동이나 경험을 요약할 수 있는 정보가 된다. 3명의 사용자로부터 수집한 데이터로 제안한 방법의 성능을 평가해 본 결과, 직관적으로 중요한 경험에 관계된 개념을 얻을 수 있었다.

1. 서 론

모바일 기기는 그 특성상 사용자가 늘 가지고 다니며, 그에 따라 모바일 기기를 이용하여 사용자 정보를 수집하려는 시도가 많이 이루어지고 있다[1]. 수집된 사용자 정보는 사용자에게 맞는 서비스를 제공하거나 사용자가 이전에 있었던 일을 기억하는 데 도움을 주는 방향으로 이용될 수 있다. 하지만 수집되는 사용자 데이터의 양이 방대하기 때문에 수집된 데이터를 모두 이용하는 것보다는 수집된 데이터 중에서 사용자에게 의미 있는 행동이나 경험들을 선별하여 서비스 제공이나 사용자 분석에 이용하는 것이 효과적일 것이다.

이전의 연구에서는 스마트 폰을 이용하여 사용자의 데이터를 수집하고 수집된 데이터를 바탕으로 키그래프를 생성하여 사용자가 수행한 행동 가운데 핵심적인 이벤트를 선택하고자 하였다[6]. 이를 통해 특정 MP3 음악 듣거나 전화통화 혹은 SMS 문자 메시지가 중요하다라는 것을 파악할 수는 있었지만 MP3 음악을 듣거나 전화통화를 한 행동이 왜 중요한 지 혹은 그렇게 중요하다고 나타난 이벤트를 통해서 사용자가 어떤 행동을 했는지는 직관적으로 이해하기 어려웠다.

본 논문에서는 스마트 폰에서 수집한 GPS 위치 정보를 장소명으로 변환하여 컨셉넷에 입력하여 직관성이 높은 정보로 변환한다. 컨셉넷은 사람이 일상생활에서 의식하지 않고 사용하는 상식들을 의미망의 형태로 구축해 놓은 것으로 컨셉넷에 특

정 단어를 입력하면 그 단어와 관련된 개념들을 단어나 어휘의 형태로 0에서 1사이의 값을 가지는 연관도와 함께 반환해준다. 컨셉넷에서 얻은 개념들을 이용하여 문서를 만들고 그 문서에 키그래프를 적용하여 가장 핵심이 되는 개념을 찾아낸다. 찾아낸 개념들은 사용자가 하루 동안 행한 일들을 요약하는 핵심적인 정보가 된다. 이 개념들은 효과적인 사용자 행동 분석이나 서비스에 이용될 수 있다. 그림 1은 GPS 데이터로부터 핵심 개념을 추출하는 시스템의 흐름을 나타내고 있다.

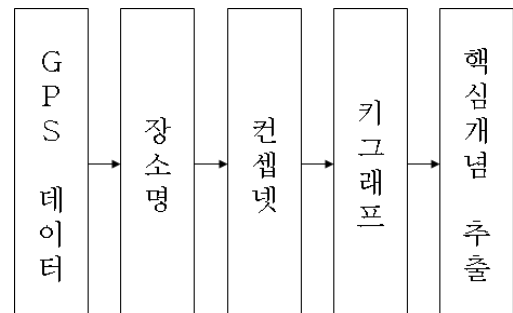


그림 1. 시스템 구성

2. 관련 연구

2.1 컨셉넷 (ConceptNet)

사람은 대화를 하거나 글을 쓸 때, 혹은 글을 읽을 때 글이나

대화 자체에서는 나타나지 않는 지식을 사용하는 경우가 많이 있다. 문맥상에서는 나타나지 않지만 실제로 사람들은 이미 알고 있는 배경지식을 가지고 글이나 대화에 나타나는 단어의 뜻이나 사실을 오해하지 않고 이해할 수 있다. 그러나 컴퓨터가 사람이 쓴 글을 이해할 때, 혹은 사람의 대화를 이해하려고 할 때 이러한 일반 상식을 가지고 있지 않기 때문에 여러 가지 문제가 발생할 수 있다.

컨셉넷은 컴퓨터가 이런 일반상식에 관련된 문제를 처리할 수 있도록 하는 시도의 하나이다. 컨셉넷은 일반 사용자들 약 1만 4천명이 Open Mind Common Sense 웹 사이트에 입력한 지식을 가지고 만들어졌으며, 70만개의 영어 문장, 30만개의 개념과 160만개의 사실(assertion)을 포함하고 있는 의미 지식 베이스이다[2]. 컨셉넷 이전에도 Cyc, WordNet 같은 시도가 있었으나 Cyc의 경우에는 CycL이라는 특수한 언어를 통해서만 사용할 수 있어 일반적인 응용 프로그램에 적용하는데 어려움이 있었고, WordNet의 경우에는 사용방법은 비교적 간단하지만 단어 사이의 관계를 표현하는 종류의 개수가 적고, 단어의 양도 컨셉넷에 비하여 적었다. 컨셉넷은 WordNet 만큼 이용이 쉬우면서도 Cyc 만큼 많은 데이터를 포함하고 있으며 단순히 계층적 분류형식의 정보가 아닌 실제로 대화나 글에서 이용되는 상식들을 포함하고 있다. 본 논문에서는 사용자가 하루에 방문한 장소들과 관련된 개념을 얻어 내기 위해서 컨셉넷을 이용하고 있다.

2.2 키그래프

키그래프는 본래 문서를 요약해주는 기능을 가지고 있는 알고리즘이다. 초기에는 문서에서 핵심적인 키워드를 찾아내주는 방안으로 개발되었다[4]. 키그래프는 문서를 작성한 저자가 그 문서에서 나타내고자 하는 가장 근본적인 개념들이 문서 내부에 있는 단어들로 표현된다고 가정한다. 그렇게 가정한 문서의 모든 단어와 문장들은 저자가 표현하려는 가장 핵심적인 단어들을 나타내기 위해서 보조하는 역할을 하며, 실제로 나타내려고 하는 것들은 단 몇 개의 단어로 요약될 수 있다.

키그래프는 기본적으로 문서에 특정 단어가 나타나는 빈도를 이용하여 핵심 단어를 찾아내지만 단순히 어떤 단어가 나타나는 빈도수가 높다고 해서 핵심 단어로 선택하지는 않으며 발생 빈도가 낮더라도 중요한 단어를 핵심 단어로 선택할 수 있다.

키그래프의 장점은 특별히 학습을 하거나 휴리스틱이 주어지지 않더라도 연속적으로 반복 입력되는 데이터 속에서 발생빈도와 사건들의 연결 관계를 가지고 어느 것이 중요한 사건인지 판단해준다는 것이다. 따라서 문서가 아니라 할지라도 정보가 문서형식으로 정리되어 주어지면 키그래프를 적용하여 중요한 정보를 찾는데 응용할 수 있다. 이런 특성을 이용하여 키그래프는 과거의 지진과 진앙 데이터를 통해서 미래에 큰 지진이 발생할 가능성이 있는 진앙을 예측하거나[3] 컴퓨터를 이용하는 사용자의 파일 이용 기록을 이용하여 많은 파일 가운데 어떤 파일이 사용자에게 중요한 의미를 가지는 파일인지 찾아내기 위해 이용되었다[5].

3. 일상생활 요약

본 논문에서는 사용자의 하루를 요약하는 개념을 찾아주기 위해서 키그래프와 컨셉넷을 이용한다. 키그래프와 컨셉넷에 사용되는 사용자 데이터는 스마트폰에 연결된 GPS 장치로부터 수집한다. 그림 2는 사용자가 방문한 장소명을 컨셉넷에 입력하고 컨셉넷에서 얻은 개념들로 생성한 문서에 키그래프를 적용하여 핵심 개념을 추출하는 방법을 보여주고 있다.

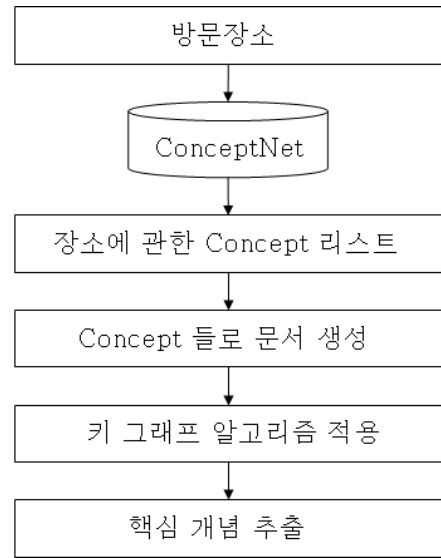


그림 2. 컨셉넷과 키그래프를 이용한 핵심 개념 추출

3.1 GPS 데이터의 레이블링

본 논문에서는 스마트폰에서 수집된 GPS 데이터를 이용한다. GPS 데이터는 위도와 경도를 표시하기 때문에 실제로 사용자가 방문한 장소를 알아내기 위해서는 위도, 경도 좌표를 레이블링된 장소명으로 변경해주어야 한다. 그림 3은 그 과정을 나타내고 있다.

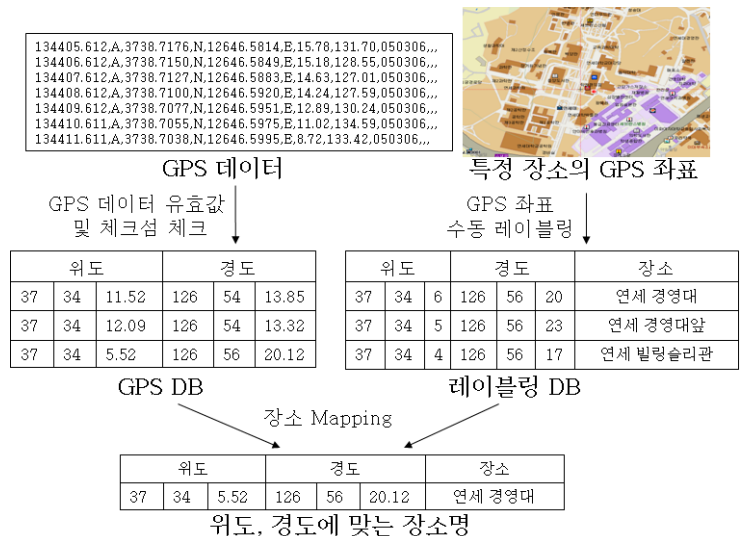


그림 3. GPS 데이터의 레이블링 및 장소명 변환

3.2 방문 장소의 영문화 및 컨셉넷에의 입력

사용자의 위도, 경도 좌표 값을 가지고 레이블링된 데이터를 이용하여 방문한 장소명을 알아냈으면 장소명을 컨셉넷에 입력하여 관련된 개념을 얻는다. 그림 4는 장소명을 한글에서 영문으로 변경하고 컨셉넷에 입력하는 과정을 보여준다.

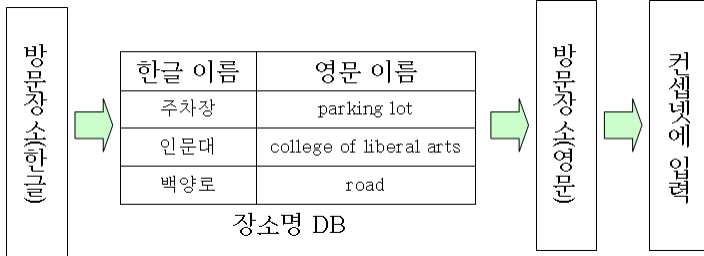


그림 4. 장소명을 영문 변환 후 ConceptNet에 입력

3.3 컨셉넷에서 반환된 값을 바탕으로 문서 작성

컨셉넷에서 장소명에 대해서 반환되는 개념들을 가지고 키그래프에서 이용될 문서를 생성한다. 컨셉넷에서 반환되는 개념의 개수가 많기 때문에 키그래프를 구성하는 개념의 개수를 적절한 수준으로 한정하기 위해서 반환된 개념들 중 연관도가 0.01 이상인 개념들만 문장을 구성하는 데 이용한다. 그림 5는 특정 장소에 대해서 컨셉넷이 반환하는 개념의 예를 보여준다.

입력 장소: parking lot
 반환되는 개념: parking lot, driveway, something, wait, keep, vegetation, concrete, parking space, yellow school bus, asphalt, parking, pavement, schoolbus,

그림 5. 특정 장소 (parking lot)에 대해 반환되는 값

문서를 생성할 때 중요한 것이 문장(sentence)을 정의하는 부분이다. 키그래프 알고리즘의 특성상 문장을 나누는 것에 따라 개념 사이의 연결 관계와 연결 강도가 결정되기 때문에 키그래프의 성능에 큰 영향을 주게 된다.

텍스트 문서에서는 마침표(.)가 나타날 때마다 문장을 나눌 수 있지만 연속되는 사건인 사용자의 GPS 데이터는 그렇게 나눌 수가 없다. 본 논문에서는 사용자가 방문하는 장소가 변경될 때마다 장소에 따라서 사용자의 행동과 업무가 달라졌다고 본다. 그러므로 특정 장소를 컨셉넷에 입력했을 때 반환되는 모든 개념들이 하나의 문장에 포함된다고 본다. 즉, 방문 장소마다 반환되는 모든 개념이 하나의 문장을 형성하게 된다.

그림 6은 실제로 생성된 문서의 예제를 보여주고 있다. 각 개념은 #기호로 구분되며, 각 문장은 마침표로 구분된다.

cross#construction#cross street#
 two woman#dreamcatcher#
 highway#grave#.

그림 6. 생성된 문서의 예

3.4. 키그래프 적용

1) 발생 빈도가 높은 개념과 연결 관계의 추출

컨셉넷에서 반환된 개념들로 생성된 문서에서 발생 빈도가 높은 개념들을 추출하고 각 문장에서 개념들이 함께 나타나는 빈도를 계산하여 각 개념의 연결 강도를 계산한다.

개념들의 발생 빈도 $|x|_s$ 는 문장(Sentence) S 에서 특정 개념 x 가 나타나는 빈도를 의미한다. $assoc(w_i, w_j)$ 는 개념 w_i 와 w_j 가 함께 나타나는 빈도를 나타내며, 특정 문장 S 에서 w_i 와 w_j 가 각각 나타나는 빈도를 계산하여 그 중에서 작은 것을 선택한다. 모든 문장에 대해서 얻어진 빈도의 합이 두 개념 사이의 연결강도를 나타낸다. 실제 계산식은 (1) 과 같다.

$$assoc(w_i, w_j) = \sum_{s \in D} \min(|w_i|_s, |w_j|_s) \quad (1)$$

각 개념의 발생빈도와 연결강도를 가지고 그림 7과 같은 그래프를 만들어 낼 수 있다.

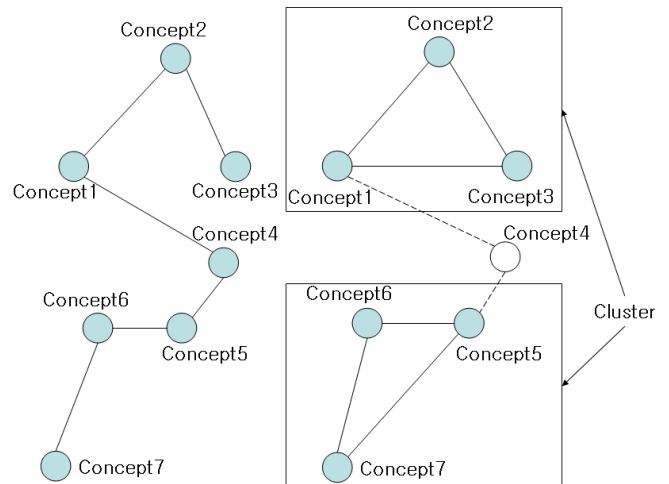


그림 7. 발생빈도와 개념 사이의 연결 강도에 따른 그래프와 클러스터의 형성

그래프를 형성하는 개념은 발생빈도가 가장 높은 개념들 가운데서 선택한다. 선택되는 개념의 개수는 실험자가 경험적으로 선택한다. 그래프의 개념들을 연결하는 개념 사이의 관계는 연결 강도($assoc$)가 강한 것부터 선택한다. 선택되는 연결의 개수는 항상 (선택된 개념의 개수 - 1) 개로 한다. 연결의 개수가 개념의 개수보다 항상 한 개 적기 때문에 모든 노드들끼리 연결 관계가 형성된다면 그림 7의 왼쪽에서 보이는 것처럼 하나로 연결된 그래프가 생성된다. 그러나 하나의 개념에서 다른 개념에 도달할 수 있는 경로가 두 개 이상 생기면 그래프에 나타나는 개념들은 모두 연결될 수 없고 두 개 이상의 덩어리로 나뉘게 된다. 이렇게 나뉜 각 부분을 클러스터라고 한다.

2) 클러스터들을 연결하는 개념을 추출

비록 빈도수는 낮더라도 문서에서 중요한 개념들이 존재한다. 그림 7의 오른쪽 그림에서 Concept4는 비록 발생빈도가 낮더라도 각각의 클러스터들을 연결해주는 개념이므로 각각의 개념 클러스터가 Concept4를 나타내기 위한 근거로 사용되었다고 가정할 수 있다. Concept4는 발생빈도와 관계없이 중요한 개념으로 취급될 수 있다. 따라서 문서의 개념들에 대해 각 개념과 클러스터 사이의 관계를 계산하여 클러스터들과의 연관관계가 높은 개념들을 추출한다.

$$based(w, g) = \sum_{s \in D} |w|_s |g - w|_s \quad (2)$$

$$neighbors(g) = \sum_{s \in D} \sum_{w \in s} |w|_s |g - w|_s \quad (3)$$

$$|g - w|_s = \begin{cases} |g|_s - |w|_s & \text{if } w \in g, \\ |g|_s & \text{if } w \notin g. \end{cases} \quad (4)$$

$$key(w) = probability(w \cap g) \quad (5)$$

$$key(w) = probability(\bigcup_{g \in G} (w|g)) \quad (6)$$

$$key(w) = 1 - \prod_{g \in G} (1 - \frac{based(w, g)}{neighbors(g)}) \quad (7)$$

수식 (2)의 *based*는 특정 개념 *w*가 클러스터 *g*와 가지는 관계의 연결 강도이며, 수식 (3)의 *neighbors*는 문서에 있는 모든 개념과 클러스터 *g*가 가지는 연결 강도의 합이다. *based*와 *neighbors*를 구하면 문서에서 특정 개념과 특정 클러스터가 가지는 관계의 연결 강도를 0과 1사이의 확률 값으로 구할 수 있다. 이 값을 수식 (5)에서 *key*라 하며, *key* 값이 큰 순서대로 실험자가 설정한 개수만큼 개념을 추출한다. *key* 값이 큰 개념일수록 클러스터들을 연결하는 개념일 가능성이 높다.

3) 문서를 요약하는 키워드의 추출

발생 빈도가 높은 개념 리스트(High Frequency Concept)와 클러스터를 연결하는 확률이 높은 개념 리스트(Key Concept)를 통하여 문서를 요약하는 키워드를 추출한다.

추출하는 방법은 앞서 추출된 개념 리스트에 다음 공식을 적용하여 큰 값을 가지는 개념들을 추출한다.

w_i : *key* 값이 큰 개념(High Frequency Concept)

w_j : 발생 빈도가 높은 개념(Key Concept)

$$column(w_i, w_j) = \sum_{s \in D} \min(|w_i|_s, |w_j|_s) \quad (8)$$

결국 위의 *column*값이 높은 개념 쌍들이 클러스터를 연결하는 중요 키워드로 선택되며, 따라서 문서를 요약하는 키워드가 된다.

4. 실험 및 결과

본 실험에서는 3명의 여대생들에게 약 2주에 걸쳐 수집한 GPS 데이터를 바탕으로 실험을 진행하였다. GPS 데이터 자체에 오류가 있는 것이나 데이터 체크섬에 이상이 있는 레코드들을 제외한 총 20만 개 이상의 레코드들을 대상으로 논문의 알고리즘을 적용하였다.

4.1 GPS 레이블링 및 영문 장소명 맵핑

먼저 사용자의 위도, 경도 좌표를 장소명으로 변환하기 위하여 레이블링을 수행하였다. 사용자가 이동했던 모든 장소를 레이블링 하는 것은 물리적으로 불가능하기 때문에 사용자가 가장 자주 이동한 장소 105곳을 레이블링하여 사용하였다.

각각의 장소명은 컨셉넷에 입력으로 이용하기 위해서 영문으로 변환되어야 한다. 여기서는 사용자의 방문 빈도가 높은 순서대로 총 105개의 국문 장소명을 영문 장소명으로 변환하여 컨셉넷에 입력한다. 그림 8은 GPS 데이터를 레이블링 장소에 대해서 처리하고 한글 장소명과 영문 변환명을 함께 나타낸 프로그램의 화면이다.

시작시간	종료시간	장소명(한글)	장소명(영문)
2006-03-02 09:10:03	2006-03-02 09:10:03	학교앞 횡단보도	crossing
2006-03-02 09:10:21	2006-03-02 09:13:14	연대앞 버스 정류장	bus stop
2006-03-02 09:17:14	2006-03-02 09:17:17	도로	road
2006-03-02 10:44:29	2006-03-02 10:45:38	백양로	road
2006-03-02 10:45:39	2006-03-02 10:45:08	학생회관앞	square
2006-03-02 10:55:40	2006-03-02 10:55:51	학생회관앞	square
2006-03-02 10:55:52	2006-03-02 11:01:17	백양로	road
2006-03-02 10:56:05	2006-03-02 10:56:20	도서관앞	garden
2006-03-02 15:05:45	2006-03-02 15:05:51	새관앞 버스 정류장	bus stop
2006-03-02 15:06:09	2006-03-02 15:06:10	백양로	road
2006-03-02 15:06:14	2006-03-02 15:06:14	학교앞 횡단보도	crossing
2006-03-02 15:06:18	2006-03-02 15:08:39	연대앞 버스 정류장	bus stop
2006-03-02 15:08:40	2006-03-02 15:09:01	공학원외부출구	exit
2006-03-02 16:24:39	2006-03-02 16:26:25	백양로	road
2006-03-02 16:26:52	2006-03-02 16:26:54	학생회관	student union building
2006-03-02 17:10:33	2006-03-02 17:10:53	도서관	library
2006-03-02 18:15:13	2006-03-02 18:15:20	구룡대앞	garden
2006-03-02 18:15:16	2006-03-02 18:16:18	주차장	parking lot
2006-03-02 18:16:19	2006-03-02 18:19:17	구룡대	college of engineering
2006-03-02 18:19:41	2006-03-02 18:21:00	공학원	Yonsei engineering research center
2006-03-02 18:21:37	2006-03-02 18:26:13	연대앞 버스 정류장	bus stop
2006-03-02 18:26:21	2006-03-02 18:26:22	학교앞 횡단보도	crossing

그림 8. GPS 데이터를 통한 방문 장소 및 영문 변환

4.2 키그래프를 통한 문서의 핵심 개념 추출

본 논문에서는 키그래프를 통하여 문서에서 추출되는 핵심 개념의 개수를 6개로 한정하고 그래프로 표현되는 연결과 개념들의 숫자를 18개로 한정하였다. 그 이유는 하나의 장소명에 대해서 컨셉넷에서 반환되는 개념의 숫자가 평균 427개나 되기 때문에 관련된 연결과 개념들을 모두 나타낼 경우 키그래프의 복잡성이 커지고 그 결과 실험자가 분석하기 어려워지기 때문이다.

표 1. 3월 4일의 사용자 스케줄

시간	분	장소	감정	행동
12	00	집	피곤	기상
13				
14				
15				
16	30	버스 정류장	신남	인사동으로
17	10	인사동	신남	도착. 맛집에서 식사
18	30	버스 정류장	신남	대학로로 이동
19	00	대학로		고속터미널로 이동
20				
21				
22	20	서산	피곤	도착
23	00	집(서산)	피곤	도착. 취침

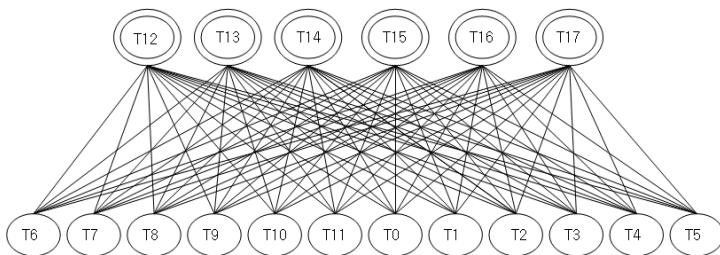
그림 9는 사용자에게 얻은 GPS 데이터 중 3월 4일에 대해서 키그래프를 생성하여 얻은 핵심 개념이다. curb, automobile, highway, alley, cable, concrete 의 6개 개념이 추출 되었다.

표 1은 3월 4일의 사용자 보고서로부터 얻은 스케줄을 나타내고 있다. 표 1에서 보는 바와 같이 사용자는 그날 약속으로 인한 이동이 잦았으며, 특히 서산의 고향집에 내려가기 위해서 자동차로 오랜 시간 이동했음을 알 수 있다. 따라서 핵심 개념으로 뽑힌 curb, automobile, highway 등은 도로를 이용한 이동이 많았던 날의 중요한 키워드가 될 수 있다.

그림 10은 3월 4일에 대해서 생성된 키그래프를 그래프 형태로 나타낸 것이다. 그림 10의 각 노드 ID에 해당되는 개념은 아래에 나타나 있다. 그림 10에서 두 개의 테두리를 가진 원으로 표현된 개념들이 추출된 핵심 개념들이며, 표에서는 색상으로 표시되어 있다.

KeyEvent	Description
T12	curb
T13	automobile
T14	highway
T15	alley
T16	cable
T17	concrete

그림 9. 3월 4일 사용자의 행동을 요약하는 핵심 개념



ID	개념	ID	개념
T0	road	T9	cross
T1	freeway	T10	asphalt
T2	tread	T11	traffic
T3	red station wagon	T12	curb
T4	intersection	T13	automobile
T5	carry	T14	highway
T6	shrubbery	T15	alley
T7	national interstate highway	T16	cable
T8	overpass	T17	concrete

그림 10. 3월 4일에 대해 생성된 키그래프

4.3 이전 연구의 결과와 비교

표 2. 같은 데이터에 대한 이전 연구와의 결과 비교

사람	날짜	핵심 개념 (현재 연구)	핵심 이벤트 (이전 연구)	사건
H1	3/4	curb, automobile, highway, alley, cable, concrete	Call54, Call32, Call53, Call55, Call56, Call17, Call2, Call57, Call1, Call3, Call81, Call5	인사동, 대학로, 고속버스로 고향집 이동
H1	3/9	road, freeway, tread, intersection, carry, shrubbery	MM18, MM19, MM12, MM26, MM10, MM11, MM27, MM25, MM15, MM16, MM17, MM20	청송대 산책, 대공연
H1	3/17	find, tread, intersection, carry, shrubbery, national interstate highway	결과없음	진달래길 걷기, 청송대 산책, 과외
H1	3/21	expressway, take walk, green, watch film, sit on chair, relaxation	결과없음	야외실습 수업

표 2는 스마트폰에서 수집된 같은 데이터를 대상으로 이전 연구에서 제안한 방법과 본 논문에서 제안한 방법을 적용한 결과를 보여주고 있다. 이전 연구에서는 MM(음악파일), Call(전화), SMS(문자), Pic(사진)의 4개 종류의 사용자 이벤트에 대해서 어떤 것이 중요한 이벤트인지를 알려줄 수 있다. 이러한 핵심 이벤트들은 사용자에게 대해 파악하는 데 중요한 정보가 된다. 본 연구에서는 특정한 핵심 이벤트를 검출해내기보다는 사용자

가 하루 동안 방문한 장소에 관련된 개념을 추출하는 것으로 보다 직관적으로 사용자의 행동을 이해할 수 있도록 하였다.

5. 결론 및 향후 연구

본 논문에서는 스마트폰을 이용하여 사용자의 GPS 데이터를 수집하고 수집된 데이터를 컨셉넷을 이용하여 관련된 개념들로 변환하였다. 그리고 그 개념들에 대해서 키그래프를 적용하여 사용자의 하루를 요약할 수 있는 핵심 개념을 추출하였다. 실험자는 추출된 개념을 통해 사용자가 하루 동안 수행한 중요한 행동을 직관적으로 이해할 수 있다. 향후 연구로는 더 나은 성능을 얻기 위해서 컨셉넷에서 관련된 어휘를 추출할 때 연관도 뿐만 아니라 연관 관계의 종류에 따라 구분하여 부정적인 관계를 배제하고 긍정적인 관계의 개념만을 고려해볼 것이다. 또한 사용자가 요약된 개념들을 살펴보고 개념을 선택할 수 있도록 한다면 정확도를 향상 시키고 사용자 평가를 더욱 수월하게 진행하는 데 도움이 될 수 있을 것이다.

참고문헌

- [1] M. Raento, A. Oulasvirta, R. Petit, and H. Toivonen, "ContextPhone: A prototyping platform for context-aware mobile applications," *IEEE Pervasive Computing*, vol. 4, no. 2, pp. 51-59, 2005.
- [2] H. Liu, and P. Singh, "Conceptnet: a practical commonsense reasoning tool-kit," *BT Technology Journal*, vol. 22, no. 4, pp. 211-226, 2004.
- [3] Y. Ohsawa, "Keygraph as risk explorer in earthquake-sequence," *Journal of Contingencies and Crisis Management*, vol. 10, pp. 119-128, 2002.
- [4] Y. Ohsawa, N. E. Benson and M. Yachida, "Keygraph: Automatic indexing by co-occurrence graph based on building construction metaphor," *Proc. Advanced Digital Library Conference (IEEE ADL'98)*, pp. 12-18, 1998.
- [5] Y. Ohsawa, "Get timely files from visualized structure of your working history," *Knowledge-Based Intelligent Information Engineering Systems*, pp. 546-549, 1999.
- [6] M.-C. Jung and S.-B. Cho, "User's information management based on keygraph for smartphone," *Proceedings of KFIS Autumn Conference*, vol. 15, no. 2, pp. 199-202, 2005.
- [7] H. Liu, & P. Singh, "Commonsense Reasoning in and over Natural Language," *KES'2004*, pp. 293-306, 2004.
- [8] K. Panu, M. Jani, K. Juha, K. Heikki, and M. Esko-Juhani, "Managing Context Information in Mobile Devices," *IEEE Pervasive Computing*, vol. 2, no. 3, pp. 42-51, ISSN 15361268, July-September 2003.
- [9] J. Curtis, D. Baxter, J. Cabral, "On the Application of the Cyc Ontology to Word Sense Disambiguation," *the Nineteenth International FLAIRS Conference*, pp. 652-657, May 2006.
- [10] G. Miller, "WordNet: An Electronic Lexical Database," May 1998