

목적을 가진 게임을 이용한 사용자의 검색 의도 자동 획득 시스템 설계

양영욱*, 임희석*

*고려대학교 컴퓨터교육학과

e-mail : brilliant_yyw@korea.ac.kr

The automatic acquisition system design of the user's searching intention using Game With A Purpose(GWAP)

Yeong-Wook Yang*, Heui-seok Lim*

*Dept of Computer Science Education, Korea University

요 약

정보의 양이 기하급수적으로 증가함에 따라, 그 정보들을 효과적으로 제공하기 위해 검색 기술들이 개발 되었다. 현대 검색 기술은 단순 텍스트 매칭 기반으로 사용자의 의도와는 상관없는 정보까지 검색하는 경우가 많아서, 실제로 원하는 정보를 찾는데 많은 시간과 노력을 요구한다. 이런 단점들을 극복하고자 사용자의 검색 의도를 고려한 검색 기술들이 개발되었다.

검색 의도를 획득하는 방법에는 자동적인 방법과 수동적인 방법, 그리고 반자동적인 방법이 있다. 자동적인 방법은 사람이 아닌 컴퓨터를 통해서 획득이 이루어지기 때문에 사용자의 의도를 완벽히 반영하는 것이 어렵다. 수동적인 방법은 사람이 직접 참여하여 작성하기 때문에 사용자의 의도를 반영할 수 있지만, 시간과 비용이 많이 드는 단점이 있다. 반자동적인 방법은 위의 두 가지 단점을 보완할 수 있다.

본 논문에서는 반자동적인 방법인 목적을 가진 게임을 이용한 사용자의 검색 의도 획득 시스템을 사용한다. 이 후에 시스템에서 도입한 Game With A Purpose(GWAP)의 개념과 시스템의 구성, 그리고 발전방향 및 결론을 논의한다.

1. 서론

인터넷의 급속한 발전으로 정보의 양이 기하급수적으로 증가하였다. 이 정보들을 효과적으로 제공하기 위해 검색 기술들이 개발되었다.

현대 검색 기술은 단순 텍스트 매칭 기반의 검색으로 사용자가 입력한 질의어에 대해서 데이터를 제공하는 형식이다. 이런 검색 기술은 사용자의 의도와는 상관없는 정보까지 검색하는 경우가 많아서, 실제로 원하는 정보를 찾는데 많은 시간과 노력을 요구한다는 단점이 있다. 이런 단점을 극복하기 위해 사용자의 검색 의도를 고려한 검색 기술들이 개발되었다[1].

사용자의 검색 의도를 반영하기 위해서는 검색 의도를 획득해야 한다. 검색 의도를 획득하는 방법에는 자동적인 방법과 수동적인 방법, 그리고 반자동

적인 방법이 있다. 자동적인 방법은 사람이 관여하지 않기 때문에 사람의 모든 생각을 반영할 수 없다는 단점을 가지고 있다. 수동적인 방법은 사람이 직접 하는 것이기 때문에 사람이 생각하는 것들을 반영할 수 있다. 하지만 자동적인 방법에 비해서 많은 시간과 비용이 요구된다. 반자동적인 방법은 자동적인 방법과 수동적인 방법의 중간 형태로서 사람이 일정부분 관여하는 방법이다. 즉, 자동적인 방법과 수동적인 방법에 단점을 보완할 수 있다.

본 논문에서 자동적인 방법과 수동적인 방법에 단점을 보완할 수 있는 반자동적인 방법을 사용한다. 본 논문에서 제안하는 반자동적인 방법은 목적을 가진 게임(Game With A Purpose (GWAP))을 이용한 검색 의도 획득 방법이다.

2. 관련 연구

2.1 Game With A Purpose(GWAP)

GWAP은 컴퓨터를 이용하는 시간 중 게임에 투자하는 시간을 AI 알고리즘을 학습시키고 컴퓨터 문제를 해결하기 위한 방법을 찾는 데서 나왔다[2].

GWAP은 게임을 만들어서 사람이 게임에서 재미를 느끼며 게임이 가진 목적을 달성한다. 그 목적은 사람이 직관적으로 쉽게 처리할 수 있지만 컴퓨터가 쉽게 못하는 일이다[3].

GWAP을 디자인 할 때는 고려해야 될 몇 가지 사항들이 있다. 우선 사람의 능동적인 참여를 유발하기 위해 게임은 재미있어야 하고 쉬워야 한다. 그리고 게임에서 느낄 수 있는 재미를 증폭시켜주기 위해서 부가적인 성취감을 느낄 수 있도록 만들어 주어야 한다. 예를 들자면 게임에서 얻을 수 있는 포인트나 레벨, 게임 상에서의 순위 등이 있을 수 있다. 의미의 사용자와 게임을 플레이해야 하고, 게임 도중에 대화는 배제되어야 한다. 이런 제약은 게임을 하는 것에 집중하는 것이 아니고, 단순히 부가적인 것을 더 챙기기 위해서 플레이어들 간에 사전모의를 배제하기 위한 것이다. 이 외에도 여러 가지 고려해야 될 사항들이 있다. 이러한 것들은 데이터의 질을 높이는 데 중요한 역할을 한다[2].

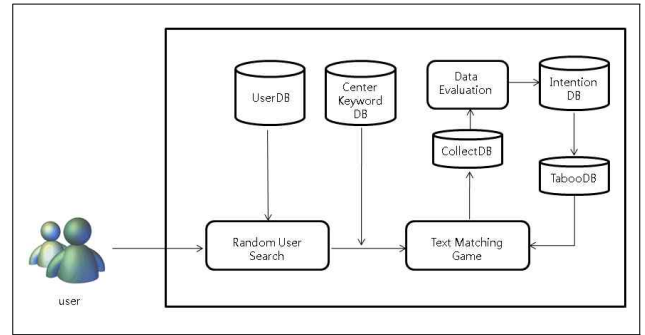
GWAP의 성공적인 사례를 예로 들어보면 ESP 게임을 들 수 있다[4][5]. ESP 게임 방법은 사용자 간에 이미지를 보고 연상되는 단어들을 매치시키는 게임이다. ESP게임에 목적은 구글의 이미지(image)를 사람이 직관적으로 분류해서 레이블(label)을 붙이는 것이다. 이런 작업으로 분류한 이미지가 100만개에 이른다[3].

위에 사례에서 볼 수 있듯이 이미지를 컴퓨터로 처리하기 위해서는 어렵다. 하지만 사람은 직관적으로 이미지에 대해서 처리할 수 있다.

본 논문에서는 컴퓨터로 획득하기 어려운 사용자들의 검색의도를 획득하기 위해서 GWAP의 개념을 이용한다.

3. 목적을 가진 게임을 이용한 검색 의도 획득 시스템

[그림 1]은 시스템에 전체적인 구성도이다. 이후에 시스템의 각 모듈과 데이터의 질을 높이기 위한 방법들에 대해서 논의한다.



[그림 1] 시스템 구조

3.1 임의의 사용자 검색

임의의 사용자 검색 모듈은 사는 지역이나 IP를 통해서 최대한 관련이 없는 사람들을 찾는다. 사용자와 관련된 사람이 게임을 같이 하게 된다면, 시스템에서 제공하는 부가적인 것들을 목적에 두고 서로 사전에 모의해서 게임에 참여할 수 있다. 이러한 경우에는 쓸모없는 데이터가 모일 수 있고, 모인 데이터의 질도 떨어진다.

3.2 단어 매칭 게임

본 시스템에서 제시하고자 하는 게임은 같이 게임하는 사용자들에게 동일한 중심 단어를 보여준다. 각 사용자들은 그 중심 단어를 보고 자유 연상을 통해서 그 단어와 연관된 단어들을 제시한다. 제시한 단어가 동일하다면 수집 데이터베이스에 들어간다. 수집된 데이터는 사용자들이 직접적으로 한 단어에 대해서 떠올린 것들이기 때문에 사용자 의도라고 볼 수 있다. 하지만 여기서 수집된 데이터가 사용자들에 조작에 의해서 만들어진 데이터가 될 수도 있기 때문에 데이터에 대한 검증이 필요하다.

3.3 데이터 평가

저장된 데이터는 사용자의 의도를 반영하지만 확실한 데이터인지 평가가 필요하다. 데이터 평가 기준을 측정하는 가장 쉬운 방법은 일정한 범위를 정하는 것이다. 즉 제시된 단어의 횟수가 일정한 범위를 초과한다면 그 데이터는 검증된 데이터라고 보는 방법이다.

검증된 데이터는 의도 데이터베이스에 저장된다. 의도 데이터베이스 안에 있는 정보가 사용자의 의도를 획득한 정보이다. 또한 검증된 데이터에 대해서 사용자들이 다시 제시를 하지 못하도록 금지어 데이터베이스에 저장된다.

3.4. 데이터의 질 확장

시스템 구조에서는 없지만 더 의미 있는 데이터를 얻기 위해서는 사용자가 게임에 더 집중할 수 있는 동기를 부여해야 한다. 본 시스템에서 제시하는 동기 부여 방법은 아래와 같다.

3.4.1 포인트와 랭킹 도입

포인트와 랭킹을 도입하는 것은 하나의 동기부여가 된다. 사용자들은 자신이 보유하고 있는 포인트를 증가하기 위해서 능동적으로 참여를 하게 된다. 또한 다른 사람의 생각을 고려하게 됨으로 집중력을 가지고 게임에 임한다.

또한 사용자의 포인트로 하여금 높은 점수를 보유하고 있는 사용자 목록을 보여준다면, 자신이 그 목록에 올라가기 위해서 집중할 것이다. 또한 그 목록에 올라가 있는 사용자는 그 목록에 계속 남기 위해서 게임에 더 능동적으로 참여를 할 수 있다[2][6].

3.4.2 반복적 데이터

같은 데이터를 가지고 계속적으로 하다보면, 자신의 생각이 아닌 형식적으로 대답을 하게 된다. 그렇다면 더 좋은 데이터를 얻는 것은 어렵다. 일정한 대답만 나오기 때문에 새로운 출력물이 만들어지지 않기 때문이다. 이러한 문제를 해결하기 위해서는 한 사용자에게 노출되는 단어에 대해 반복 기준을 정해두고 플레이를 해야 한다[2][6].

3.4.3 출력물 제약

위에 데이터 평가 부분에서 언급했던 것처럼 너무 일반적이거나 많이 나온 출력에 대해서 입력할 수 없도록 제제를 해야 된다. 그래야 새로운 출력물을 유도할 수 있고, 더 높은 질의 데이터를 얻을 수 있다[2][6].

3.4.4 플레이어 테스트

사용자가 입력한 데이터가 확실한 데이터인지를 테스트 하는 것이다. 랜덤한 방법을 사용했다고 하더라도, 사용자들이 높은 포인트를 위해서 확실하지 않은 데이터를 입력했다면 그것도 제제를 가해야 되기 때문이다[2].

4. 향후 발전 방향 및 결론

이 시스템에 대해 개발과 실험을 하지 않았기 때문에 구현과 실험이 필요하다. 이후에 더 좋은 시스템을 개발하기 위해서는 게임에 대한 획기적인 아이디어를 계속적으로 생각해야 한다. 사람의 능동적인 참여가 많아야 빠르고 더 정확하게 사용자의 의도를 획득할 수 있기 때문이다. 또한 데이터를 확실하게 검증할 수 있는 방법을 연구해야한다. 단순히 제시된 횟수로서 데이터를 검증하는 것보다 수치화된 해석이 필요하다.

참고문헌

- [1] 허선영, “온톨로지 기반의 시맨틱 검색 시스템에 대한 연구”, 한국정보처리학회 춘계학술발표대회 논문집, 제14권 제1호, pp. 463-466, 5월, 2007.
- [2] Luis von Ahn and Laura Dabbish “Designing Games With A Purpose”, August, 2008.
- [3] Ling-Jyh Chen and Bo-Chun Wang, Kuan-Ta Chen, “Playing GWAP With Strategies”, February, 2009.
- [4] GWAP, www.gwap.com
- [5] Luis von ahn, “Game with a Purpose”, June, 2006
- [6] Barbora Hladk’a and Jiř’ı M’ırovsk’y and Pavel Schlesinger, “Designing a Language Game for Collecting Coreference Annotation”, August, 2009.