

스마트 미디어 검색 기술 동향 및 전망

조준면, 김명은, 홍진우
한국전자통신연구원

요약

스마트 미디어 환경은 과거 PC 기반의 정보 또는 콘텐츠 생산/유통/소비 환경과 비교하여 다층적이고 서로 연관된 새로운 다양한 특성을 내포하고 있는데, 본고에서는 이러한 스마트 미디어 환경에서 정보 검색 기술의 동향과 향후 발전 방향을 알아본다. 우선, 기존 검색 시스템을 검색 방식과 검색 형태 측면에서 살펴본 후, 검색 서비스에 영향을 미치는 스마트 미디어 환경의 특성 또는 변화 유발요인을 정리한다. 이후 이러한 변화요인들에 대처하기 위해 예상되는 기술적 혁신 방향과 각 발전 방향의 구체적인 내용을 알아본다.

I. 서론

근래의 스마트 미디어 혁명은 비즈니스 영역에서 뿐만 아니라 연구개발 영역 등 다양한 영역에서 큰 충격과 변화를 야기하고 있다. 좁게 보면, 스마트 미디어 환경은 스마트폰, 스마트패드 등과 같은 스마트 기기에 의해 촉발되었고 따라서 이러한 스마트 기기를 이용한 정보 또는 콘텐츠의 생산/유통/소비 환경이라 정의할 수도 있으나, 넓게 보면, 모바일 컴퓨팅 (mobile computing), 소셜 컴퓨팅 (social computing), 개인화 컴퓨팅 (personalized computing) 등 새로운 컴퓨팅 패러다임에 의해 촉발되었고 따라서 이러한 컴퓨팅 기법을 이용한 정보 또는 콘텐츠의 생산/유통/소비 환경이며 이 환경에서의 주요 매체가 스마트 기기라 정의할 수 있다.

사실 스마트 기기의 확산과 새로운 컴퓨팅 패러다임의 등장은 서로 영향을 주는 양방향의 관계를 가지고 있다고 보아야 할 것이다. 인터넷 접근성과 이동성이 탁월한 스마트 기기의 등장과 확산에 따라 과거와는 비교할 수 없을 정도의 다양한 형태와 방대한 양의 정보 및 콘텐츠가 생산되고 유통/소비되고 있으며 이러한 방대한 정보 및 콘텐츠에 기반하여 보다 지능적인 서비스가 비로서 가능해 졌고 지능적인 서비스에 대한 요구가 증대되

었다. 반대로, 지능적 정보/콘텐츠 서비스가 가능해 지고 지능적 서비스에 대한 사용자 요구가 증대됨에 따라 새로운 사업모델 및 관련 기술이 발전되었고 이는 다시 스마트 기기의 확산과 보편화를 촉진하고 있다.

스마트 미디어 환경은 과거 PC 기반의 정보 또는 콘텐츠 생산/유통/소비 환경과 비교하여 다층적이고 서로 연관된 새로운 다양한 특성을 내포하고 있다. 이러한 새로운 특성에 의해 과거의 스마트 미디어 서비스 영역은 새롭게 재편되고 있으며 개별 서비스의 핵심 기능 및 개발 방향이 변화되고 있다.

본고에서는 넓은 의미의 스마트 미디어 환경 관점에서 검색 서비스 분야의 기술 동향 및 발전 방향에 대해 논한다. 먼저 기존 검색 시스템을 검색 방식과 검색 형태 측면에서 살펴본 후, 검색 서비스에 영향을 미치는 스마트 미디어 환경의 특성 또는 변화 유발요인을 정리한다. 이후 이러한 변화요인들을 대처할 수 있는 향후 검색 관련 기술적 혁신 방향을 논하고 각 발전 방향의 구체적인 내용을 알아본다.

II. 검색 시스템의 분류

1. 검색 방식에 따른 분류

현재 검색 시스템은 검색 방식 측면에서 크롤러 기반, 디렉토리 기반, 하이브리드 검색, 메타 검색 방식으로 구분할 수 있다.

클로러 기반 방식의 검색 시스템에서는 스파이더, 클롤러, 웹 봇 등으로 불리는 자동화된 에이전트 프로그램을 이용하여 웹 상의 문서를 자신의 데이터베이스에 다운로드하고 저장한다. 사용자의 검색 요청은 검색 키워드를 저장된 웹 문서의 인덱스에서 찾아 해당 문서의 링크를 제공함으로써 처리된다. 이 방식은 지속적인 웹 문서 수집 및 점진적인 확장이 용이하여 방대한 범위의 정보 검색이 가능한 장점을 가진다. 구글 검색 시스템이 대표적인 예이다.

디렉토리 기반 방식의 검색 시스템에서는 사람에 의해 웹 사이트들이 사전에 정의된 특정 디렉토리에 분류 저장되고, 저장된

웹사이트들이 사전에 정의된 규칙에 의해 랭킹된다. 사용자의 검색 요청은 키워드 매칭에 의해 찾아진 웹 문서를 디렉토리별로 그룹핑하여 제공함으로써 처리된다. 이 방식은 정보의 분류에 사람이 개입함으로써 사용자가 정확한 정보를 찾을 수 있도록 한다는 장점을 가진다. 야후, 네이버 검색 시스템이 대표적 예이다.

하이브리드 방식의 검색 시스템에서는 클롤러 방식과 디렉토리 방식을 병용하며 일반적으로 사용자에게 더 좋은 검색결과를 제공한다. 이 방식은 이미지, 비디오, 웹 페이지, 책 등 다양한 형태의 검색 결과를 제공하는데 장점을 가지며, 스폰서 페이지 리스트를 통해 추가적인 정보를 사용자에게 제공한다. 야후, 네이버 검색 시스템도 하이브리드 방식이라 볼 수 있으며 마이크로소프트의 MSN검색이 대표적 예이다.

메타 검색 방식의 시스템에서는 다른 검색 시스템의 검색 알고리즘과 평가 기준을 활용한다. 즉, 다른 검색 시스템의 검색 결과를 병합하여 사용자에게 제공한다. 다양한 관점 및 기준의 검색 결과를 통합적으로 제공함으로써 사용자의 정보 요구를 최대한 충족시킬 수 있다는 장점을 가진다. Metacrawler 시스템이 대표적인 예이다.

2. 검색 형태에 따른 분류

검색 시스템은 검색 형태에 따라 다시 다중차원 검색, 패시트 검색, 클러스터 검색, 자연어 검색, 소셜 검색, 시맨틱 검색 등으로 구분할 수 있다.

다중차원 검색에서는 다양한 검색 방식을 혼용하며, 사용자 검색 요구에 관련된 유용한 결과를 최대한 많이 식별하여 제공하기 위해 다중 검색 결과에 대해 다양한 정제 및 추론을 적용하는 후처리 과정을 거친다.

패시트 검색은 수작업으로 구성된 계층구조의 분류체계를 통해 검색결과를 제공하고 사용자가 계층적 분류체계를 통해 관련 없는 결과를 배제하거나 검토 범위를 좁혀 나감으로써 자신이 원하는 정확한 정보를 획득할 수 있도록 하는 탐사적 재질의 방식의 확장 형태라 볼 수 있다.

클러스터 검색에서는 검색 결과가 주요 주제에 따라 그룹핑된 형태로 제공된다. 그룹은 동적으로 구성될 수 있으며 한 그룹에 속한 검색 결과들은 서로 밀접한 관계를 가진다.

자연어 검색에서는 사용자의 의도에 부합된 검색결과를 제공하기 위해 자연어 형태로 입력된 사용자 검색문의 구문분석을 통해 언어적/의미적 정보를 검색에 최대한 이용한다.

소셜 검색에서는 사용자의 소셜 네트워크 정보를 이용한다. 소셜 네트워크 정보는 검색 인터페이스에 이용될 수도 있고, 검색 범위 및 검색 결과의 선정/정제에 이용될 수도 있다.

시맨틱 검색에서는 검색 결과의 정확도를 높이기 위해 사용자의 의도 및 검색문의 의미를 해석한다. 사용자 검색문의 의미는 시맨틱 웹 (Semantic Web) 기술을 바탕으로 자연어 처리, 논리 추론, 어휘 사전, 텍스트 마이닝 등 다양한 기술을 활용한다.

Ⅲ. 검색기술의 변화 요인

스마트 미디어 환경에서 새로운 검색 기술을 유인하는 요인은 크게, 인터넷 접근의 용이성, 소셜 네트워크/미디어의 확산, 다양한 모바일 애플리케이션의 등장, 온라인 데이터의 폭발적 증가, 사용자 요구사항의 변화 등으로 요약할 수 있다.

1. 인터넷 접근의 용이성

스마트폰 등 이동형 스마트 기기의 보편화와 무선 인터넷 액세스 포인트의 확산에 따라 사용자의 인터넷 접근이 점점 쉽고 단순해 지고 있으며, 나아가 사용자가 의식하지 못하는 사이 인터넷 접속 환경이 장소와 시간의 구애 없이 개인 사용자에게 제공되고 있다. 따라서 사용자는 과거와는 달리 다양한 목적과 대상에 대해 검색 서비스를 언제 어디서나 쉽게 요청할 수 있게 되었다. 이러한 인터넷 접근의 용이성은 기존 검색 시스템으로 하여금 다양한 기기 또는 매체를 통해 다양한 형태의 정보 및 콘텐츠에 대한 검색 결과를 제공할 수 있는 기능의 확장을 요구하고 있다[1]. 즉, 모바일 단말 사용자의 상시 휴대성을 감안하여 언제, 어디서나 사용자의 검색 요구에 대해 정보를 제공해야 하고, 상대적으로 작은 화면을 고려하여 사용자가 쉽게 정보를 습득할 수 있도록 검색 결과를 맞춤형으로 제공해야 한다[2].

2. 소셜 네트워크 및 소셜 미디어의 확산

블로그, 트위터 등 다양한 형태의 소셜 네트워킹 플랫폼 (즉, 소셜 네트워크 사이트)의 등장과 소셜 네트워크를 기반으로한 다양한 정보/콘텐츠의 폭발적인 증가에 따라 검색 시스템은 보다 우수한 검색 성능을 요구 받고 있다 [1]. 즉, 검색 시스템은 일반 검색의 효율성을 높이기 위해 소셜 네트워크를 이용할 수 있다. 예를 들어, 소셜 관계의 적절성을 따져 지인의 정보 또는 지인과 관계된 정보를 더 중요시하여 검색결과를 차별적으로 제공하거나 현재 검색 도메인의 전문가 네트워크를 이용하여 좀더 신뢰성 있는 정보를 제공할 수 있다. 더 나아가 검색 시스템은 이벤트를 기반으로 구성되는 동적 소셜 그룹을 찾아내고 이러한 동적 그룹에 관련된 소셜 미디어를 검색하여 제공함으로써 신뢰성있는 정보를 제공할 수 있다[3].

3. 다양한 모바일 애플리케이션의 등장

최근 스마트 미디어 환경이 확산됨에 따라 다양한 모바일 애플리케이션이 등장하고 있다. 과거 게임 및 작업 관리 용도의 애플리케이션에서 최근의 지도/지리정보 애플리케이션, 전자상거래 애플리케이션, 기상정보 애플리케이션, 교통정보 애플리케이션 등 유용한 정보를 제공하는 다양한 애플리케이션이 가용하게 되었다. 검색 시스템은 이러한 다양한 정보를 이용하여 사용자의 선호 및 상황에 맞는 검색결과를 제공할 수 있다[1]. 예를 들어, 사용자 위치에 근접한 식당, 마트, 서점, 여행사 등을 우선순위 높게 제공할 수 있으며, 더 나아가, 실시간 교통정보를 바탕으로 가장 빠른 길을 찾아줄 수도 있다. 위치 관련 다양한 정보를 지도위에 표시하여 제공함으로써 사용자의 정보 활용을 획기적으로 증대시킬 수 있다[4].

4. 온라인 정보의 폭발적 증가

근래 각종 연구와 통계자료를 통해 확인할 수 있듯이, 온라인상의 정보는 가히 폭발적으로 증가하고 있다. 이러한 정보의 기하급수적 증가는 검색 시스템에 새로운 도전 과제와 기회를 동시에 주고 있다. 즉, 검색 시스템은 방대한 정보를 빠르게 처리할 수 있어야 하는 반면, 사용자가 원하고 사용자에게 도움이 되는 정보를 찾을 수 있는 가능성이 점점 높아지고 있다[1]. 더 나아가, 방대한 데이터의 상호 연결과 의미를 분석하여 숨은 정보를 발견함으로써 궁극적으로 사용자의 올바른 의사결정과 문제 해결을 도울 수 있다. 실제로 최근 이러한 서비스를 위한 빅데이터 분석 (big data analysis) 기술에 대한 관심이 집중되고 있다.

5. 사용자 요구사항의 변화

위에서 살펴본 검색기술의 변화 유발 요인들이 복합적으로 작용하여 결과적으로는 과거와는 다른 사용자 요구사항을 낳고 있다. 즉, 새로운 기술의 도래, 새로운 미디어 환경의 확산, 다양한 정보/콘텐츠 형태의 등장 등이 검색 시스템에 대한 사용자의 기대를 변화시키고 그 수준을 높이고 있다. 이제 사용자는 확장된 맞춤형, 개인화, 지능화 등 자신들의 검색 경험과 검색 결과의 유용성을 획기적으로 증대시킬 수 있는 새로운 특성들을 검색 시스템에 기대하고 있다.

사용자들은 자신의 의도와 부합되지 않는 정보를 점점 더 스팸으로 간주하고 있으며 따라서 검색 시스템은 방대한 정보에서 사용자에게 스팸으로 인식되는 정보를 어떻게 효과적으로 제거함으로써 성능지표를 유지할 것인가를 고민해야 한다[1]. 사용

자는 작은 화면 등 인터페이스의 제약이 있는 미디어 환경에서 PC 환경에서와는 달리 자신의 검색 요청과 관련 없는 정보가 검색 결과에 포함되는 것을 원치 않는다. 즉, 사용자의 검색요청과 관련 있는 검색결과를 틀림없이 제공하는지를 나타내는 재현율 (recall rate) 보다는 실질적으로 관련 있는 검색결과만을 제공함으로써 관련 없는 정보가 함께 포함됨으로써 사용자의 정보 습득을 방해하지 않기를 원한다 (정확도: precision rate).

사용자는 원하는 정보를 최종적으로 습득하는데 소비되는 시간에 점점 더 민감해 지고 있다. 따라서 스팸 제거, 정확도 향상과 함께 검색결과를 일목요연하게 요약 정리하여 제공하는 기능이 요구된다[1]. 실제로 최근 구글에서는 '지식 그래프' 서비스를 시작하였다. 기존에는 검색 결과를 보여줄 때 질의어와 관련된 웹페이지들의 리스트만을 보여 주다가, 지식 그래프에서는 구글이 다양한 미디어 분석을 통해 직접 수집한 정보와 이들간의 관계를 그래프 개념으로 제공함으로써 사용자가 원하는 정보를 정확하고 빠르게 찾아갈 수 있도록 하고 있다.

마지막으로 사용자는 개인화된 정보를 원한다. 검색 시스템은 사용자의 개인 정보, 선호 정보, 소셜 네트워크 등을 이용하여 검색 범위, 검색 결과 순위 등을 사용자에게 따라 다르게 조정/제한할 수 있어야 한다. 이를 통해 사용자는 검색 시간을 줄이고 검색 결과의 신뢰성을 확보할 수 있다.

IV. 기술적 혁신 동향

본 절에서는 III절에서 살펴본 검색 기술의 변화를 유인하는 다양한 요인들에 대처하기 위한 기술적 혁신 동향을 살펴본다.

기술적 변화 유인 요인들을 다시 한번 요약하면 검색 결과의 양 보다는 질과 신뢰도가 더욱더 중요해 지고 있으며, 사람 즉, 사용자 중심의 정보 세상을 지향하고 있다고 할 수 있다. 사실, 휴대 단말의 보편화에 따른 인터넷 접근의 용이성 증대, 소셜 네트워크 활성화에 따른 소셜 정보의 증가, 다양한 모바일 애플리케이션의 등장에 따른 가용 정보의 증가 등 모든 요인에서 공통적으로 정보 검색 결과의 정확도 및 신뢰도라는 요인을 발견할 수 있는 것과 같이, III절에서 살펴본 요인들은 서로 독립적이지 않고 상호 연관되어 있다.

이러한 요인들을 종합적으로 고려할 때, 이 분야의 많은 전문가들은 방대한 데이터의 상호 연결과 의미 분석을 통해 숨은 정보를 발견하고, 궁극적으로 의사 결정을 보다 쉽고 정확하게 할 수 있는 정보 검색 서비스로 발전하리라 예측하고 있다[5]. 이러한 예측에 포함된 기술적 혁신은 크게 시맨틱 검색, 소셜 검색, 개인화 검색 이라는 세 범주로 정리할 수 있다.

1. 시맨틱 검색

최근의 시맨틱 웹 기술은 웹 상의 콘텐츠에 의미와 구조를 부여함으로써 컴퓨터가 다른 정보간의 관계를 이해할 수 있고 지능적 추론을 통해 암묵적인 지식을 유도할 수 있게 하였다[6].

이러한 시맨틱 검색 기술은 이미 2000년대 초에 주요 대학을 중심으로 기초연구가 시작되었는데, 단계별로 5가지 유형으로 구분해 볼 수 있다. 의미 모호성 해소와 어휘 개념 확장 검색은 전통적으로 자연언어 처리 부문에서 관심을 가졌던 영역으로, 의미 모호성 해소는 정보 검색의 정확도 (precision rate) 향상이 주요 목표가 되며, 어휘 개념 확장의 경우 검색 재현율 (recall rate)을 향상시키는데 초점이 맞춰져 있다. 개체 특성 확장과 연관 주제 확장의 경우, 정보가 가진 특성을 구체적으로 이해하고, 관련 정보간의 상호 관계를 파악함으로써, 숨은 의미를 이해할 수 있도록 한다. 마지막으로, 최근 큰 관심을 받고 있는 의도기반 검색(intent based search)은 사용자의 검색 의도와 목적을 파악하고, 그에 적합한 정보와 연관 서비스를 추천하는 것을 목표로 한다. 의도기반 검색을 구현하기 위해서는, 사용자 의도를 올바르게 인지하고, 이를 서비스와 연결하기 할 수 있는 기술의 확보가 필수적이며, 사용자 상황 인지, 사용자 피드백 학습, 대화형 UX를 통한 의도 구체화 등의 매우 다양한 방법론이 연구되고 있다. 주목할 점은 구글, 마이크로소프트, 애플 등의 세계적 기업들이 최근 Powerset, Metaweb(Freebase), Siri 등을 인수 하고, 시맨틱 검색 관련한 내부 연구 개발 조직을 확대, 강화하여 그 사업 범위를 빠르게 확대해 나가고 있는 것이다[5].

시맨틱 검색과 기존 검색 시스템과의 가장 큰 차이점은, 인덱스의 구조에서 찾을 수 있다. 기존 검색 시스템은 형태소 분석 등의 과정을 통해 문서로부터 키워드(토큰)를 추출해 내고, 이들을 역파일(invert file) 구조의 인덱스에 저장 하게 된다. 반면에 전형적인 시맨틱 검색 시스템은 문서로부터 시맨틱 네트워크를 추출, 이를 그래프 구조로 지식베이스에 저장하고 검색 시에는 지식베이스에 질의하는 방식을 취하게 된다. 매우 다양한 관점의 시맨틱 검색 기술이 개발되고 있으나, 어떤 형태던 시맨틱 네트워크로 정보가 표현, 재구성되어 지식 베이스에 저장, 질의 하는 형태를 가진다. 시맨틱 네트워크는 개념(컨셉, 용어)과 그들간의 관계가 표현된 지식 표현 방법 중 하나이다[5].

〈그림 1〉은 사용자 관점에서의 정보 검색과 분석의 진화 단계를 보여준다. 모든 정보 시스템 구현의 시작은 정보를 생산하고 안전하게 저장하는 구조를 확보하는 것이다. 우리는 이 단계를 level 0로 정의한다. 정보가 쌓여가기 시작하면, 이를 관리하고 싶어진다. 디렉토리를 만들어 분류를 하거나, 제목, 날짜 등의

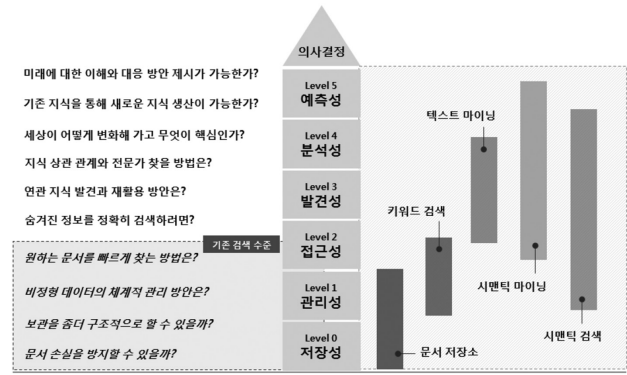


그림 1. 정보 검색 발전 단계 [5]

메타데이터로 태그를 다는 형태가 전형적이다. 초기 야후는 이 방법을 통해 시장 선점이 가능했다. 그러나 관리체계가 잘 갖춰진 경우도 문서의 수가 수백만 개 이상이 되면, 심지어 웹과 같이 수천억, 수조 개 이상이 되면, 리 체제만으로는 필요한 문서를 찾아 활용하는 것이 점점 어려워진다. 이때가 비로서 정보 검색의 중요성이 인식되는 단계인 level 2가 된다. 구글은 혁신적인 저장 및 인덱싱 구조와 성능 좋은 랭킹 알고리즘을 통해 엄청난 양의 문서로부터 사용자 만족도가 높은 검색 성능을 달성할 수 있었다. 구글은 level 2 단계에서 지배자가 된 것이다 [5].

기존 검색 시스템이 제공하는 방대한 정보에의 빠른 접근성은 상당기간 시장을 만족시켜 왔다. 그러나 이제는 검색 결과 조차 사람이 모두 읽고 이해하기 힘들 정도로 많아지게 되었고, 단지 특정 키워드를 포함하는 문서를 찾아 주는 것으로는 사용자 요구를 충족하기가 점점 힘들어 지고 있다. Level 3 발견성 확보와 level 4 분석성 제공은, 검색 결과를 정리, 요약, 분석함으로 그 안에 숨겨진 핵심 정보와 상호 연관성, 트렌드 등의 패턴을 이해하도록 돕는다. 특히, 텍스트마이닝 기술은 방대한 비정형 정보로부터 의미 있는 정보를 추출하거나 재조직화 함으로, 정보의 발견성을 더욱 용이하게 한다. 시맨틱 검색의 최종 목표는 의사 결정 지원이며, 이를 위해서는 시맨틱 웹 기술에 기계 학습 등의 마이닝 기술이 적용된 시맨틱 마이닝 기술의 확보가 관건이 될 것이다. 향후 5년 내에 시맨틱 마이닝 기술의 발전과 검색 서비스에의 결합은 실질적인 시맨틱 검색 서비스 시스템 출시를 가능하게 할 것이며, 2020년 무렵에는 데이터 지능화를 통한 예측성 수준에 이르게 될 것으로 기대된다[5].

2. 소셜 검색

최근, 사용자의 소셜 프로파일, 소셜 네트워크, 그리고 소셜 미디어 등에 기반하여 상황에 맞는 개인화된 검색결과를 제공하는 기술은 주요 검색 서비스 제공 회사의 핵심 전략이 되고

있다. 예를 들어, 2009년부터 서비스한 구글의 소셜 검색은 사용자들이 자신의 구글 계정 정보를 트위터, 유튜브, 플리커 등과 연계하면 지인들이 사이트에 올린 사진 등의 콘텐츠를 검색해 주는 것이다. 처음에는 검색 창과는 별도로 하단에 일부의 사용자에게만 제공하였으나 2011년 구글은 검색어나 관계의 적절성을 분석해 지인의 콘텐츠를 상위에 보이도록 하거나, 지인이 트위터 등을 통해 링크한 콘텐츠도 검색 결과의 상위로 검색되도록 소셜 기능을 강화했다. 즉 일반 검색 결과 창에 블로그 등에 지인이 올린 정보가 상위에 보이게 되면서 각 사용자마다 서로 다른 검색 결과를 갖게 된다[3].

이러한 소셜 검색의 기반이 되는 소셜 네트워크 관련한 연구 분야는 매우 광범위하고, 여러 가지 기술이 복합적으로 연결되어 있다. 그 중 소셜 네트워크 분석(SNA) 연구는 사회학과 인문학에서부터 시작됐다. 이러한 소셜 네트워크는 사람을 점으로 표현하고 관계를 선으로 표현하는 그래프나 관계를 나타내는 선에 weight를 뒀서 밀접도를 나타내는 그래프로 표현할 수 있다. <그림 2>는 이렇게 표현된 소셜 관계의 한 예를 나타낸다. 하단 영역의 노드는 사람을 나타내고 에지를 통해 다른 사람과 연결되어 있다. 이렇게 연결된 그래프는 관계 분석을 통하여, 가족, 친구, 팀(직장)과 같은 소셜 그룹으로 묶이게 된다. 그 위의 중간 영역은 이벤트를 기반으로 한 동적 소셜 그룹을 나타낸다. 예를 들어, '아버지 생신 가족 모임'을 보면 가족 중에 실제 행사에 참가한 가족과 친구가 그 동적 소셜의 멤버가 된다. 마지막으로 상단 영역은 그것으로 인해 만들어지는 소셜 미디어(자료)를 나타내는데, 사진이나 기록을 의미하고 인터넷이나 웹에서 공유가 가능하다[3].

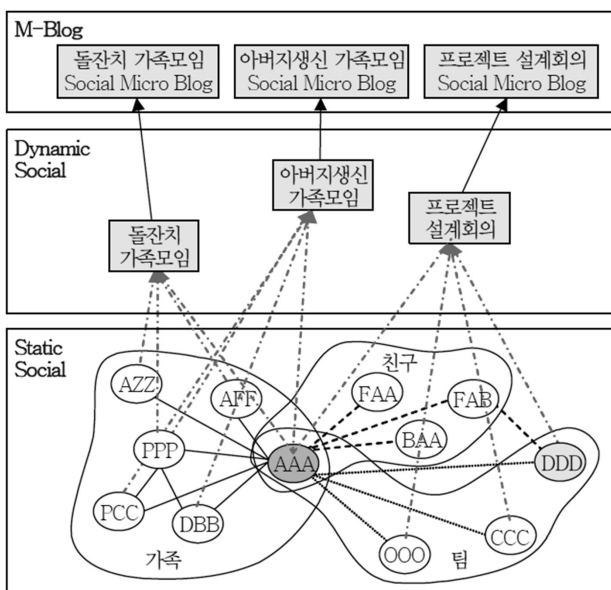


그림 2. 소셜 그래프의 예 [3]

3. 개인화 검색

앞 소셜 검색 절에서 논한 소셜 그래프를 기반으로 사용자 맞춤형 정보를 제공하는 것도 개인화 검색에 포함된다고 볼 수 있는 것처럼 개인화 검색의 경계를 명확히 설정하기는 쉽지 않다.

일단, 사용자의 관심 또는 선호, 사용자의 최근 정보 검색/소비 패턴, 사용자의 위치 등에 기반하여 검색 범위와 결과를 제한함으로써 사용자가 원하는 정보를 습득하는 시간을 줄여주는 검색을 개인화 검색이라 정의할 수 있을 것이다.

개인화 검색을 위해서는 사용자의 성향을 파악하여 사용자의 개인 정보를 모델링하여 사용자에게 의미있는 검색 결과를 제시할 수 있어야 한다. 개인화 기술은 일반적으로 그림 3과 같이 사용자의 성향을 파악하는 batch process 부분과 이러한 데이터를 이용하여 실제 사용자에게 적용하는 online process로 구분할 수 있다. Batch process를 좀더 세분하면 사용자의 개개인의 행동을 수집, 저장하는 기술, 수집된 정보를 체계화하는 프로파일링 기술, 이러한 프로파일에 기반하여 사용자의 행동을 파악하는 사용자 모델링 기술로 나눌 수 있다[2].

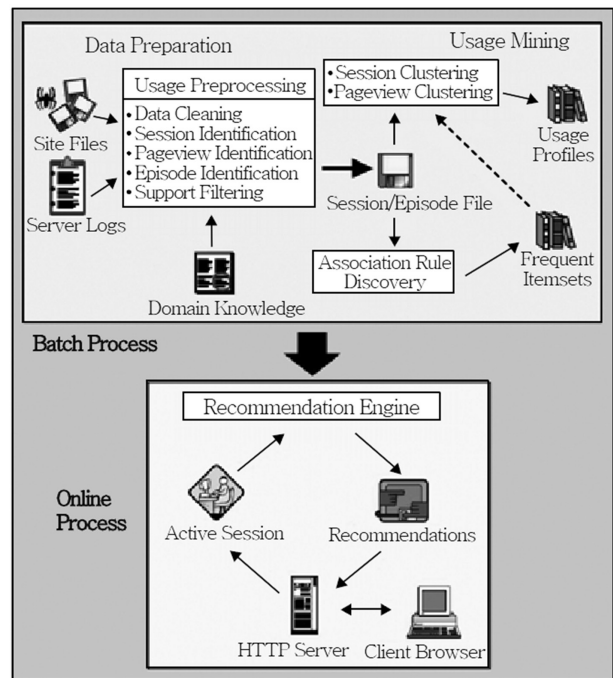


그림 3. 개인화 서비스의 일반적인 구조 [2]

실제 개인화 기술은 1990년대 말부터 2000년대 초까지 각광을 받았으나 PC를 주 사용 단말로 하는 유선 인터넷 환경에서는 사용자에게 대한 식별이 로그인을 필수적으로 해야 하기 때문에 사용자에게 외면되었고 기술이 더 이상 상용화되지 못했으나 사용자가 명확히 식별되는 모바일 환경에 적용되면서 관심이 집중되고 있으며 그 파급효과가 매우 클 것으로 예측된다.

즉, 휴대 단말은 사용자가 항상 휴대한다는 점에서 사용자의 개인 정보를 충분히 수집할 수 있는 도구가 되기 때문에 이러한 특성을 고려한 개인 맞춤형 검색 서비스에 효과적이다. 또한, 모바일 단말의 작은 화면에서 많은 검색 결과를 보기는 힘들기 때문에 모바일 사용자에게 적합한 요약된 검색 순위화가 필요하다. 이는 검색어 히스토리 및 사용자 프로파일을 통한 사용자 중심의 결과를 제시하여 해결하여야 한다[2].

검색 기술은 주어진 질의에 대해 고정된 정답을 내놓는 방식에서 개별 사용자를 고려한 결과를 제시하는 방향으로 바뀌고 있다. 위에서 언급한 검색 결과의 순위화뿐만 아니라, 사실 문서 요약(snippet), 질의어 추천 등 검색 시스템의 모든 측면이 잠재적인 개인화의 대상이 될 수 있다. 향후 사용자에 대해 충분히 많은 정보를 수집한 검색 시스템이 사용자의 질의 없이도 관련 정보를 추천하는 일도 가능할 것이다.

V. 결론

본고에서는 비즈니스 영역에서만뿐만 아니라 연구개발 영역 등 다양한 영역에서 큰 변화를 야기하고 있는 스마트 미디어 환경에서 검색 기술의 변화요인과 기술 혁신 동향을 살펴 보았다. 변화요인으로 인터넷 접근의 용이성, 소셜 네트워크의 확산, 다양한 모바일 애플리케이션의 등장, 온라인 정보의 증가, 사용자 요구사항의 변화 등에 대해 살펴 보았고, 이러한 변화 요인에 효과적으로 대처하기 위해 예상되고 있는 기술적 혁신 내용에 대해서 알아 보았다.

기술적 혁신 동향은 검색 결과의 양 보다는 질과 신뢰도를 증대할 수 있고, 나아가 사람 즉, 사용자 중심의 정보 세상을 지향하는 방향으로 전개되고 있다고 요약할 수 있다. 이러한 동향에는 웹 상의 콘텐츠에 의미와 구조를 부여함으로써 컴퓨터가 데이터의 상호 연결과 의미 분석을 통해 숨은 정보를 발견하고, 궁극적으로 의사 결정을 보다 쉽고 정확하게 할 수 있도록 지원하는 시맨틱 검색, 사용자의 소셜 프로파일, 소셜 네트워크, 그리고 소셜 미디어 등에 기반하여 상황에 맞는 검색결과를 제공하는 소셜 검색, 마지막으로 사용자의 관심 또는 선호, 사용자의 최근 정보 검색/소비 패턴, 사용자의 위치, 나아가 사용자의 의도 등에 기반하여 검색 범위와 결과를 제한함으로써 사용자가 원하는 정보를 습득하는 시간을 줄여주는 개인화 검색 기술 등이 포함된다.

가까운 시일 내에 이러한 기술적 혁신이 검색 시스템에 적용되고 높은 수준의 AI 기술과 접목됨으로써 방대한 데이터 분석과 지능화를 통한 미래 예측과 추론이 가능하게 될 것으로 예상된다.

Acknowledgement

본 연구는 방송통신위원회의 ETRI연구개발지원사업의 일환으로 수행되었음 [11921-03001, “Beyond 스마트 TV 기술 개발”]

참고 문헌

- [1] “Next Generation Search Engines - Technology Market Penetration and Roadmapping”, Frost & Sullivan, 2011.9
- [2] 임수중, 오효정, 류법모, 정호영, 장명길, “모바일 지능형 검색 기술 동향”, 전자통신동향분석 제25권 제3호, ETRI, 2010.6
- [3] 김지용, 손동환, 김현진, “소셜 네트워크 서비스 기술 동향”, 전자통신동향분석 제26권 제3호, ETRI, 2011.6
- [4] 장운섭, 오재홍, 김경옥, “지리공간 웹 기술 동향”, 전자통신동향분석 제22권 제3호, ETRI, 2007.6
- [5] 이경일, “검색의 진화, 시맨틱 검색”, White Paper, 솔트룩스, 2011
- [6] 김명은, 정인철, 조준면, “스마트 TV 검색 서비스 기술 동향”, 전자통신동향분석 제26권 제4호, ETRI, 2011.8

약 력



조 준 면

1993년 한국과학기술원 공학사
 1995년 한국과학기술원 공학석사
 2006년 한국과학기술원 공학박사
 1995년~2001년 삼성중공업 중앙연구소
 선임연구원
 2002년~현재 한국전자통신연구원 선임연구원
 관심분야: 지식베이스 시스템, 지식 표현 및 추론,
 지능형 시스템, 온톨로지



김 명 은

1996년 숭실대학교 공학사
 1998년 서강대학교 공학석사
 1998년~2000년 동양시스템하우스 SI사업부
 2000년~현재: 한국전자통신연구원 차세대스마트
 TV연구단 선임연구원
 관심분야: 스마트TV 서비스 플랫폼, 시맨틱 검색,
 소셜 미디어 검색



홍 진 우

1993년 광운대학교 대학원 전자계산기공학과
 공학박사
 2007년~2007년 성균관대 언론정보대학원
 방송통신최고위과정 수료
 1998년 독일 프라운호퍼연구소 파견연구원
 1984년~현재 한국전자통신연구원
 차세대스마트TV연구단장
 2005년~2010년 과학기술연합대학원 대학교
 겸임교수
 2008년~2008년 KAIST 겸임교수
 2012년~현재 한국정보통신학회 부회장
 관심분야: 스마트미디어, 방통융합,
 디지털방송 미디어/서비스,
 멀티미디어 프레임워크, 스마트TV