

인지 시스템으로 기업의 사고 방식 전환

IBM® Redbooks®
Point-of-View 발행물 -
IBM Academy of Technology



Rob High, IBM Fellow, Bill Rapp,
IBM Distinguished Engineer

주요 내용

인지 시스템은 기업의 사고, 행동 및 경영 방식에 전환을 가져올 수 있습니다.

- ▶ 자연어 처리, 가설 생성 및 평가, 동적 학습을 연계하여 강력하고 빠르며 정확한 솔루션을 생성할 수 있습니다.
- ▶ 비정형 데이터의 복잡성을 이해하는 데 도움이 될 수 있습니다.
- ▶ 답을 심사하고 평가하는 데 고급 분석을 적용할 수 있습니다.
- ▶ 더 똑똑하게 각 주기와 상호 작용을 수행하도록 결과에 기초한 학습 향상을 지원할 수 있습니다.
- ▶ 심층 자연어 처리(NLP)를 이용하여 무제한적인 주제에 대한 언어를 평가한 다음 정보에 기초하여 판단할 수 있습니다.

빅데이터로부터 통찰력 확보

데이터의 양이 급증하고 있습니다. 그러나 데이터의 양뿐 아니라 데이터의 형식도 갈수록 다양해지고 우리가 이러한 데이터를 이해해야 하는 속도 역시 더욱 빨라지고 있습니다. 데이터에서 최고의 가치를 얻기 위해서는 이 물밑듯 밀려오는 정보를 평가할 수 있는 컴퓨터가 필요합니다. 이러한 데이터를 **빅데이터**라고 합니다.

전 세계의 데이터 중 90%가 지난 2년간 만들어진 것입니다. 이 세상에서 상호 연결과 기능이 확대되면서 데이터의 양이 계속 증가할 것으로 예상됩니다. 그리고 전 세계의 정보 중 80%가 문헌, 보고서, 기사, 연구 논문, 이메일, 블로그, 트위터, 포럼, 채팅, 문자 메시지와 같은 비정형 정보입니다.

이 모든 데이터를 더 효과적으로 이해하고 가장 시급하고 중요한 문제의 해결에 적용할 수 있는 지식을 얻을 방법이 필요합니다. 전달되는 각종 정보를 제대로 파악하고 활용하면서 질문에 답하고 통찰력을 얻어 정보에 기초한 더 현명한 의사결정을 내릴 수 있어야 합니다. 그러나 기존의 컴퓨팅 시스템은 사람처럼 정보를 읽고 정확하게 이해하기가 쉽지 않습니다. 사람의 언어에서 발견되는 미묘한 속성, 독특한 표현, 관용적 표현과 뉘앙스를 인식할 수 있는 새로운 차원의 컴퓨팅이 필요합니다. 이를 **인지 컴퓨팅 (cognitive computing)**이라고 합니다.

인지 시스템은 사람이 추론하고 정보를 처리하는 방식을 모방합니다. 빠르게 계산하고 결정론적 작업을 수행하도록 프로그래밍된 기존의 컴퓨터와 달리 인지 시스템은 확률론적 분석 기법으로 정보를 분석하고 그로부터 통찰을 얻습니다. 데이터와의 상호 작용을 통해 학습하면서 사실상 끊임없이 스스로를 다시 프로그래밍합니다.

IBM은 인지 시스템의 시대를 여는 선두 주자입니다. 정보와 자연어를 처리하기 위해 컴퓨터와 기술을 사용하는 방식을 근본적으로 바꾸고 있으며 기업의 정보 사용 및 의사결정 방식에도 변화를 일으키려 합니다.

예를 들어, 의료 기관은 IBM Watson™ 같은 인지 시스템을 사용하여 모든 가용 데이터, 특히 텍스트 정보를 분석해 더 좋은 치료 결과를 얻고 프로세스의 효율성을 높일 수 있습니다. 금융 기관 역시 이 기술을 통해 방대한 비정형 데이터를 분석하여 여신 결정, 투자 분석, 위험 관리에 활용할 수 있습니다. 인지 시스템은 향후 기업의 사고, 행동 및 경영 방식에 전환을 가져올 수 있습니다.



Redbooks

언어 처리의 복잡성

언어는 아이디어의 표현이며 우리의 의사소통, 이해, 사고, 추측과 상상의 매체입니다. 우리가 주변 세상을 이해하고 그 이해한 바를 전달하고 다루는 방법입니다.

사람의 언어가 지니는 깊이, 복잡성과 광대함 때문에 컴퓨터가 이해하는 데 어려움이 따릅니다. 사람은 불일치, 모순, 불규칙, 격차, 모호성을 극복하면서 결국 서로의 언어를 매우 정확하게 이해하게 됩니다. 컴퓨터는 우리처럼 쉽게 盧뿔뿔할 수 없으며 사람이 추론하는 방식으로 방대한 양의 데이터를 분석하는 데 어려움을 겪습니다. 우리의 언어가 믿을 수 없을 만큼 복잡하기 때문입니다.

이와 같은 정밀도와 정확도의 차이는 중요합니다. **정밀도 (Precision)**는 하나의 구절에서 발견할 수 있는 기계적 또는 과학적 정확성입니다. **정확도 (Accuracy)**는 어떤 구절이 참으로 간주될 수 있음을 다른 구절에서 암시하는 정도입니다. 어떤 질문에 정확하게 답하기 위해서는 그 질문의 컨텍스트를 고려해야 합니다. 증거가 될 정보가 부족한 상황에서는 질문의 일부에 대해 정밀한 답을 내놓을 수 있더라도 그 질문에 정확하게 답변하기 어렵습니다. 이와 같은 정밀도와 정확도의 차이 때문에 사람의 언어를 이해하려는 컴퓨터는 한층 더 어려움을 겪게 됩니다.

사람은 때때로 혼란스럽더라도 수월하게 언어를 처리할 수 있습니다. 정보의 의미를 해석하는 능력이 컴퓨터보다 뛰어나기 때문입니다. 이러한 능력은 우리가 얻으려는 정보에 대한 컨텍스트를 제공하는 자질을 타고난 덕분입니다. 이러한 개념은 수십 년간 인공지능 분야에서 이루고자 했던 핵심 목표이기도 합니다. 언어 처리의 정밀도가 향상되었지만, 컴퓨터가 광활한 자연어의 세상에서 제 기능을 하는데 필요한 수준의 정확도에 도달한 것은 Watson 이 유일합니다.

방대한 비정형 정보를 효과적으로 탐색하기 위해서는 새로운 시대의 컴퓨팅, 즉 인지 시스템이 필요합니다.

Watson 소개

IBM Watson 은 IBM 연구소가 지난 100 여 년간 이 세상을 더 좋은 곳으로 만들고 더 똑똑한 삶의 방식을 실현할 시스템을 만들기 위해 끊임없이 노력하고 과학적 탐구에 매진하면서 이룬 기술 혁신의 대표적인 예입니다.

Watson 은 복잡성이라는 난관에도 불구하고 기술 시스템에서 패턴을 식별하고 중요한 통찰력을 얻고 의사결정을 개선하는 데 큰 발전이 있었음을 보여 줍니다. 특히 Watson 은 Deep Question-Answering (DeepQA) 기술이 진일보한 예입니다. Deep QA 는 주제의 범위가 제한되지 않은, 자연스러운 사람의 말을 이해한 다음 다뤄진 주제에 대해 정보에 근거하여 판단하는 기능을 지원합니다. Watson 에 구현된 과학 기술이 사람과 컴퓨터의 의사소통을 새로운 차원으로 발전시킬 수 있습니다. 또한 고급 분석의 기능을 확대하여 방대한 양의 정형 데이터와 비정형 데이터를 이해할 수도 있습니다. 이를테면 Watson 은 의료진이 환자를 진단하고 치료하는 데 중요하게 쓰이는 정보를 적시에 제공하기 위해 DeepQA 를 사용할 수 있습니다.

Watson 과 언어

많은 자연어 시스템에서는 구체적이고 고도로 형식화된 규칙의 테두리 내에서의 정밀도를 강조해 왔습니다. 컨텍스트에 대한 평가 없이 특정 단어와 동의어의 모음을 찾습니다. 그리고 단어를 찾아낸 총 횟수를 계산합니다. 일반적으로 이러한 개념을 **피상적 자연어 처리 (shallow natural language processing)**, 즉 **shallow NLP**라고 부릅니다. 이 방식은 좁은 범위에서 정밀할 수 있으나 컨텍스트를 고려하지 않기 때문에 정확하지 않을 때가 많습니다. 게다가 컨텍스트가 바뀌면 정밀도가 떨어질 수 있습니다.

정밀도보다 정확도가 필요할 경우, 컨텍스트를 분석하면서 질문을 평가하는 **심층 자연어 처리 (deep natural language processing)**, 즉 **심층 NLP** 기법을 사용합니다. Watson 은 가까이 있는 정보, 더 넓은 범위에서 사용 가능한 정보, **코퍼스 (corpus)**라고 하는) 지식 베이스, 소스 데이터베이스 등으로부터 최대한 많은 컨텍스트를 수집하고 평가하는 심층 NLP 시스템입니다.

인지 시스템인 Watson은 아이디어를 전달하고 다루는데 있어 사람과 비슷한 특성을 적용할 수 있습니다. 언어를 분해하여 구절 간의 추론을 찾아내고 질문에 대한 정확한 답을 구성할 수 있습니다. 인지 시스템을 기존의 컴퓨팅과 연계하여 사용한다면 방대한 정보 기반을 토대로 대규모로, 더 우수한 속도, 정확성, 복원력을 발휘하면서 문제를 해결할 수 있습니다.

그러나 Watson 역시 말 속의 단어를 반드시 이해하는 건 아니며, 그보다는 사람이 구사하는 언어적 특징을 이해합니다. 이 특징을 토대로, 상황이 변화하더라도 어떤 구절 (질문)에서 다른 구절 (답)이 추론되는지 높은 정확도로 판단할 수 있습니다.

인지 시스템이 사람과 유사한 특성으로 행동하는 데 있어 컨텍스트, 시간 및 공간 제한조건의 개념 모두 중요한 역할을 합니다.

언어 처리의 기반을 구축하려면 컨텍스트와 추론이 필수적입니다.

Watson 에서 질문에 답하는 방법

Watson 은 강력한 기능을 갖춘 전 세대 컴퓨팅을 토대로 합니다. 이 기반 기술 각각이 특별하지는 않지만 이들을 함께 사용할 경우 강력한 솔루션이 만들어질 수 있습니다.

- ▶ 자연어 처리는 비정형 데이터의 복잡성을 이해하는 데 도움이 됩니다.
- ▶ 가설 생성 및 평가는 의미 있는 증거만을 토대로 일련의 답을 심사하고 평가하기 위해 고급 분석 기술을 적용합니다.
- ▶ 동적 학습은 더 똑똑하게 각 주기와 상호 작용을 수행할 수 있도록 결과에 기초한 학습 향상을 지원합니다.

Watson 의 작동 방식을 구성하는 중요한 요소 중 하나인 코퍼스는 교재, 지침서, 입문서, FAQ, 복리후생 프로그램, 뉴스와 같은 각종 비정형 지식으로 이루어져 있습니다. Watson 은 전체 코퍼스를 수용하고 콘텐츠를 수정하거나 선별하여 사용 가능한 형식으로 만듭니다. 코퍼스 콘텐츠의 적합성 여부에 초점을 맞추고 오래되었거나 관련 없거나 출처의 신빙성이 떨어지는 정보를 걸러냅니다.

Watson 은 질문을 분해하고 코퍼스에서 가능한 답을 찾아낸 다음 그 답과 문장 컨텍스트를 수백 가지로 조사합니다. 그 결과를 활용하면서 높은 수준의 신뢰도로 질문과 잠재적 답을 해석합니다. $\pm 3\sigma$ 1 은 Watson 에서 어떤 질문에 대답하기 위해 거치는 프로세스를 보여 줍니다.

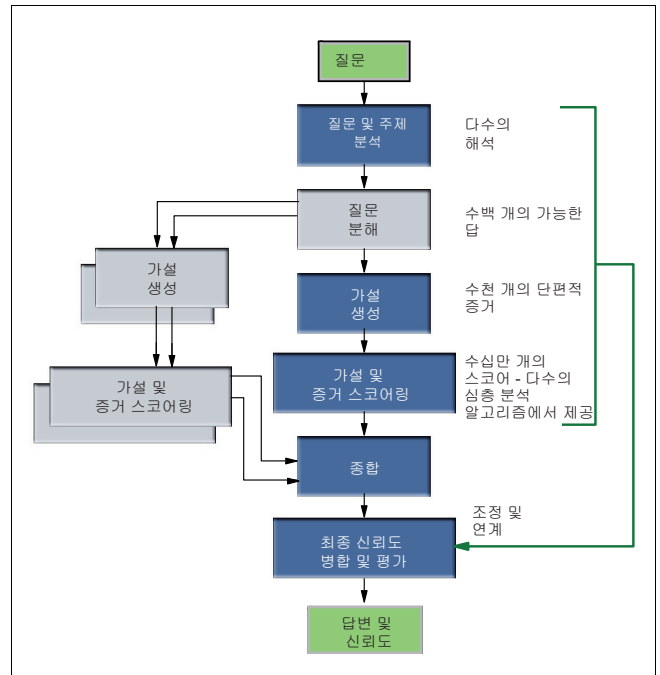


그림 1 Watson 에서 질문에 답하는 방법

Watson 이 갖춘 기능의 조합으로 다음 목표에 부합하는 솔루션이 생성됩니다.

- ▶ 프로그래밍 컴퓨팅의 제약 극복
- ▶ 정형화된 로컬 데이터에 의존하지 않고 전 세계의 비정형 데이터 심층 활용
- ▶ 의사결정 트리 기반의 결정론적 애플리케이션에서 벗어나 사용자와 함께 진화하는 확률론적 시스템 지향
- ▶ 답의 유무가 불확실한 위치의 목록을 제공하는 키워드 기반 검색 대신 신뢰도를 기준으로 답을 찾아내는 직관적이고 대화적인 방식 지향

Watson 활용 사례

각종 기업과 기관에서 기존 컴퓨팅과 함께 Watson 을 활용하면서 통계 분석, 규칙 및 비즈니스 프로세싱, 협업, 리포팅을 통해 비즈니스 문제를 해결하고 있습니다. 이를테면 투자 위험 신호가 될 만한 이벤트에 대한 Watson 의 답과 통계 분석을 연계할 수 있습니다.

의료

누구나 담당 의사가 자신의 건강 문제에 대해 정확한 최상의 답을 알고 있기를 바랍니다. 그러나 생물학 데이터가 7년마다 2배로 늘어나는 가운데 과연 의료진이 건강 증진에 필요한 모든 것을 항상 파악할 수 있는 방법은 무엇입니까?

Watson은 의료 업계에서 환자의 건강 증진에 필요한 방대한 생물학 정보를 처리, 저장, 검색하고 분석하는 데 도움이 될 수 있습니다.

Watson이 미묘한 속성을 지닌 의학 진단을 대체하지는 않습니다. 그러나 환자의 복잡한 상태를 분석하고 의료진에게 그 환자에 적합한 치료 가능성을 놀랄 만한 속도와 정확성으로 제시하여 선택의 범위를 좁힘으로써 올바른 치료 결정을 이끌 수 있습니다. 예를 들어, 가슴 통증을 호소하는 45세 환자가 있을 경우 Watson으로 환자의 상태와 병력을 모두 취합하고 분석함으로써 그 사람에게 적합한 치료법을 마련하고 더 빨리 효과를 거둘 수 있습니다.

비즈니스와 금융

인지 시스템은 기업에서 고객 및 직원과 상호 작용하는 방식에 큰 변화를 가져올 것입니다. Watson을 활용하여 기업의 위험 관리 및 평가 프로세스를 개선할 수 있습니다. 신뢰도가 높고 놀랄 만큼 빠른 시스템이 있고 여기에 올바른 질문을 한 다음 자신 있게 정답을 선택할 수 있다면 귀사의 비즈니스 환경을 어떻게 바꾸시겠습니까? 고객의 구매 패턴과 습관을 파악할 수 있다면 제품 라인과 소매 프로세스를 어떻게 바꾸시겠습니까?

금융 업계에서는 데이터의 양이 매년 70%씩 증가하고 있습니다. 이와 같이 거대하게 쌓인 데이터로부터 어떻게 신속하고 확실하게 진정한 지식과 핵심적인 통찰력을 이끌어낼 것입니까? 이용 가능한 금융 및 비즈니스 뉴스, 경제 동향, 시장 정보 등을 모두 동시에 조명함으로써 각종 정보들 사이에서 유익한 논리적 연결 고리를 찾아낼 수 있습니다. 이를테면 그러한 정보를 분석하여 조만간 인수 대상으로 삼을 만한 기업들을 탐색할 수 있습니다.

고객 서비스

시장에 Watson을 적용하는 첫 단계는 바로 콜 센터에서 이루어지며, Watson의 기능은 콜 센터 비즈니스 모델을 바꿔 놓을 것입니다.

여러 업계에서 콜 센터는 상당한 비용이 투입되는 분야이기도 합니다. 상당수의 콜 센터가 수백만 통의 전화를 받느라 대기 시간이 길어지고 그로 인해 고객의 불만이 심화되고 있습니다. Watson의 목표는 QoS (quality of service)를 강화하고 신속하고 더 정확하게

올바른 답을 제공하는 것입니다. Watson은 훨씬 더 효율적인 콜 센터 모델을 실현하고 사용자가 각자의 방식으로 데이터를 수집할 수 있게 하는, 인지 시스템 진화의 다음 단계를 여는 기술입니다.

다음 단계 : IBM의 지원

IBM은 의료, 금융, 콜 센터, 정부 공공 기관, 에너지 산업 분야에서 또 다른 혁신을 이루어 더 똑똑한 세상을 건설하는 데 Watson이 큰 역할을 할 수 있도록 적극 투자하고 있습니다.

IBM은 Watson이 앞으로도 여러 분야에서 전문가를 돕는, 더 똑똑하고 가볍고 신속한 조연자가 될 수 있도록 계속 투자할 것입니다. IBM Watson에 대한 자세한 내용은 다음을 참조하십시오.

ibm.com/innovation/us/watson

자세한 정보

이 문서에서 주로 다룬 개념에 대한 자세한 내용은 다음 자료를 참조하십시오.

- ▶ *인지 시스템의 시대 : IBM Watson과 그 작동 방식* 심층 탐구, REDP-4955

<http://www.redbooks.ibm.com/abstracts/redp4955.html?Open>

- ▶ 새로운 컴퓨팅 시대를 여는 IBM Watson

ibm.com/innovation/us/watson

- ▶ IBM Watson: 똑똑한 세상을 위한 똑똑한 답

ibm.com/innovation/us/watson/watson-for-a-smarter-planet/smarter-answers-for-a-smarter-planet.html

주의사항

이 정보는 미국에서 제공되는 제품과 서비스를 대상으로 개발된 것입니다.

IBM은 이 문서에서 언급된 제품, 서비스 또는 기능을 다른 국가에서 제공하지 않을 수도 있습니다. 한국에서 사용 가능한 제품 및 서비스에 대해서는 한국 IBM 담당자에게 문의하십시오. IBM 제품, 프로그램 또는 서비스를 언급했다고 해서 해당 IBM 제품, 프로그램 또는 서비스만을 사용할 수 있다는 것을 의미하지는 않습니다. IBM의 지적 재산권을 침해하지 않고 기능상 동등한 제품, 프로그램 또는 서비스를 대신 사용할 수 있습니다. 그러나 비 IBM 제품, 프로그램 또는 서비스의 운영에 대한 평가 및 검증은 사용자의 책임입니다.

IBM은 이 문서에서 다루고 있는 특정 내용에 대해 특허를 보유하고 있거나 현재 특허 출원 중일 수 있습니다. 이 문서를 제공한다고 해서 특허에 대한 라이선스까지 부여하는 것은 아닙니다. 라이선스에 대한 의문사항은 다음으로 문의하십시오.

135-700 서울특별시 강남구 도곡동 467-12 군인공제회관빌딩 한국 아이.비.엘 주식회사

다음 단락은 현지법과 상충하는 영국이나 기타 국가에서는 적용되지 않습니다. IBM은 타인의 권리 침해, 상품성 및 특정 목적에의 적합성에 대한 묵시적 보증을 포함하여 (단, 이에 한하지 않음) 묵시적이든 명시적이든 어떠한 종류의 보증 없이 이 발행물을 "현상대로" 제공합니다. 일부 국가에서는 특정 거래에서 명시적 또는 묵시적 보증의 면책사항을 허용하지 않으므로, 이 사항이 적용되지 않을 수도 있습니다.

이 정보에는 기술적으로 부정확한 내용이나 인쇄상의 오류가 있을 수 있습니다. 이 정보는 주기적으로 변경되며, 변경된 사항은 최신판에 통합됩니다. IBM은 이 발행물에서 설명한 제품 및 / 또는 프로그램을 사전 통지 없이 언제든지 개선 및 / 또는 변경할 수 있습니다.

이 정보에서 언급되는 비 IBM의 웹 사이트는 단지 편의상 제공된 것으로, 어떤 방식으로든 이들 웹 사이트를 옹호하고자 하는 것은 아닙니다. 해당 웹 사이트의 자료는 본 IBM 제품 자료의 일부가 아니므로 해당 웹 사이트 사용으로 인한 위험은 사용자 본인이 감수해야 합니다.

IBM은 귀하의 권리를 침해하지 않는 범위 내에서 적절하다고 생각하는 방식으로 귀하가 제공한 정보를 사용하거나 배포할 수 있습니다.

비 IBM 제품에 관한 정보는 해당 제품의 공급업체, 공개 자료 또는 기타 범용 소스로부터 얻은 것입니다. IBM에서는 이러한 제품들을 테스트하지 않았으므로, 비 IBM 제품과 관련된 성능, 호환성, 기타 주장의 정확성을 확인할 수 없습니다. 비 IBM 제품의 성능에 대한 의문사항은 해당 제품의 공급업체에 문의하십시오.

이 정보에는 일상의 비즈니스 운영에서 사용되는 자료 및 보고서에 대한 예제가 들어 있습니다. 이들 예제에는 개념을 가능한 완벽하게 설명하기 위하여 개인, 회사, 상표 및 제품의 이름이 사용될 수 있습니다. 이들 이름은 모두 가공의 것이며 실제 기업의 이름 및 주소와 유사하더라도 이는 전적으로 우연입니다.

본 문서에 포함된 모든 성능 데이터는 제한된 환경에서 산출된 것입니다. 따라서 다른 운영 환경에서 얻어진 결과는 상당히 다를 수 있습니다. 일부 성능은 개발 단계의 시스템에서 측정되었을 수 있으므로 이러한 측정치가 일반적으로 사용되고 있는 시스템에서도 동일하게 나타날 것이라고는 보증할 수 없습니다. 또한 일부 성능은 추정을 통해 추측되었을 수도 있으므로 실제 결과는 달라질 수 있습니다. 이 문서의 사용자는 해당 데이터를 본인의 특정 환경에서 검증해야 합니다.

저작권 라이선스 :

이 정보에는 여러 운영 플랫폼에서의 프로그래밍 기법을 보여주는 원어로 된 샘플 응용프로그램이 들어 있습니다. 귀하는 이러한 샘플 프로그램의 작성 기준이 된 운영 플랫폼의 응용프로그램 프로그래밍 인터페이스 (API) 에 부합하는 응용프로그램을 개발, 사용, 판매 또는 배포할 목적으로 추가 비용 없이 이들 샘플 프로그램을 어떠한 형태로든 복사, 수정 및 배포할 수 있습니다. 이러한 샘플 프로그램은 모든 조건하에서 완전히 테스트된 것은 아닙니다. 따라서 IBM은 이들 샘플 프로그램의 신뢰성, 서비스 가능성 또는 기능을 보증하거나 진술하지 않습니다.

이 REDP-4961-00 문서는 April 1, 2013 에 작성되거나 업데이트되었습니다.



상표

IBM, IBM 로고, ibm.com 은 미국 또는 기타 국가에서 사용되는 International Business Machines Corporation 의 상표 또는 등록 상표입니다. 이와 함께 기타 IBM 상표가 기재된 용어가 상표 기호 (® 또는 ™) 와 함께 이 정보에 처음 표시된 경우, 이와 같은 기호는 이 정보를 발행할 때 미국에서 IBM 이 소유한 등록상표 또는 일반 법적 상표입니다. 또한 이러한 상표는 기타 국가에서 등록상표 또는 일반 법적 상표입니다. 현재 IBM 상표 목록은 다음 사이트에 있습니다. ibm.com/legal/copytrade.shtml

다음 용어는 미국 또는 기타 국가에서 사용되는 International Business Machines Corporation 의 상표입니다.

IBM®
IBM Watson™
Redbooks®
Redbooks(logo) 

다음 용어는 타사의 상표입니다.

기타 회사, 제품 및 서비스 이름은 타사의 상표 또는 서비스표입니다.