

## 13. AWS AI 서비스

### 3강. 고급 텍스트 분석 : Amazon Comprehend

#### 학습목표

- Amazon Comprehend 개요, 주요 기능에 대해 설명할 수 있다.
- Amazon Comprehend를 이용하여 텍스트에서 감정 분석을 할 수 있다.

#### 학습내용

- Amazon Comprehend 개요
- Amazon Comprehend 주요 기능
- 실습. Comprehend를 이용하여 텍스트에서 감정 분석

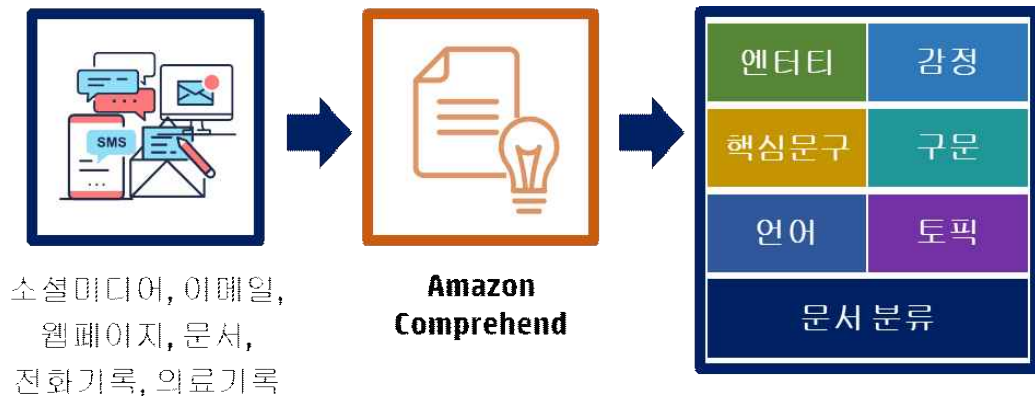
#### ■ 세상을 잇(IT)다!

- 비정형 데이터의 가치
  - ✓ 비정형 데이터에 보물이 숨겨져 있을지도 모른다?
    - 비정형 데이터(고객의 이메일, 지원 티켓, 제품 후기, 소셜 비디오, 광고 등)에는 고객의 심리에 대한 통찰이 드러나 있음
    - 이를 투입하여 사업상 효과를 얻을 수 있으나, 이를 입수할 방법이 문제
    - ML은 방대한 텍스트 모음에서 특정한 관심 항목을 가려내는 데 특히 효과적이며, 언어 뒤에 숨겨진 심리를 거의 무한대로 학습할 수 있음

#### 1. Amazon Comprehend 개요

- Amazon Comprehend
  - ✓ “Comprehend = (충분히)이해하다”
  - ✓ 머신러닝을 사용하여 텍스트 안에 있는 통찰력과 관계(핵심 문구, 장소, 사람, 브랜드 또는 이벤트 등)를 찾아내는 자연어 처리(NLP) 서비스

- Amazon Comprehend 작동 방식



- Amazon Comprehend 장점



- 텍스트에서 더 나은 답을 확보
  - ✓ Amazon Comprehend는 고객 지원 사례, 제품 리뷰, 소셜 미디어 피드, 뉴스 기사, 문서 및 기타 소스의 텍스트에서 의미와 관계를 찾을 수 있음
  - ✓ 예를 들어 고객이 제품에 만족하거나 만족하지 못할 때, 가장 자주 언급되는 기능을 파악할 수 있음
- 주제별 문서 구성
  - ✓ Amazon Comprehend는 문서 및 기타 텍스트 파일(소셜 미디어 게시물 등) 모음을 분석하고 자동으로 관련 용어 또는 주제별로 정리할 수 있음
  - ✓ 정리된 정보를 고객에게 개인화된 콘텐츠로 제공하거나 더 풍부한 검색 및 안내를 제공할 수 있음

- 자체 데이터에 대한 훈련 모델
  - ✓ Amazon Comprehend는 손쉽게 확장하여 정책 번호나 부품 코드 등 특정한 용어를 찾아낼 수 있음
  - ✓ 요청별 고객 지원 문의나 제품별 소셜 미디어 게시물 등 해당 비즈니스에 알맞은 방식으로 문서와 메시지를 분류하도록 Comprehend를 확장할 수 있음
- 일반 및 업계 특정 텍스트 지원
  - ✓ 최첨단 머신러닝 모델로 작동하는 Amazon Comprehend는 소셜 미디어 게시물, 이메일 및 웹 페이지와 같은 구조화되지 않은 텍스트에서 통찰력을 발견할 수 있음

## 2. Amazon Comprehend 주요 기능



- 핵심 문구 추출
  - ✓ 핵심 문구 추출 API는 핵심 문구 또는 논점과 이것이 핵심 문구임을 뒷받침하는 신뢰도 점수를 반환

**샘플 텍스트:** 저는 열성적인 사진 작가로서, 취미용으로 들고 다니는 DSLR이나 즉석 필름 카메라로 사진을 찍곤 합니다. 성능과 편의성은 DSLR이 최고이지만 즉석 필름 카메라에는 마술처럼 신비한 면이 있습니다. 실제 필름에 사진을 찍기 때문이거나, 촬영하는 모든 사진이 하나뿐인 물리적 실체로 남기 때문이겠죠(온갖 사진이 넘쳐나는 Instagram과 Facebook 세상에서 더욱 각별한 느낌을 주죠). 한 가지 확실한 것은, 사용하는 과정 자체가 매우 재미있고 파티에서 사람들의 눈길을 끌 수 있다는 것입니다.

[출처] <https://aws.amazon.com>

핵심 문구	개수	신뢰도
열성적인 사진작가	1	0.99
DSLR	2	0.97
즉석 필름 카메라	2	0.99
취미용	1	0.99
성능과 편의성	1	0.94
실제 필름	1	0.99
모든 사진	1	0.92
하나뿐인 물리적 실체	1	0.99
오늘	1	0.91
세상	1	0.99
Instagram과 Facebook	1	0.99

- 감성 분석

- ✓ 텍스트(긍정, 부정, 중립 또는 혼합)의 전체적인 감성을 반환

**샘플 텍스트:** 스몰 사이즈를 주문하면 딱 맞을 줄 알았는데 미디엄-라지 사이즈에 가까운 것 같아요. 품질은 매우 좋아요. 사진보다 밝은 브라운 컬러이긴 하지만 거의 비슷하고요. 내피를 면이나 울 소재로 했으면 10배는 더 좋았을 것 같아요.

[출처] <https://aws.amazon.com>

감성	점수
혼합	0.89
긍정	0.09
부정	0.01
보통	0.00

- 구문 분석
  - ✓ 토큰화 및 품사(PoS)를 사용하여 텍스트를 분석하고 텍스트 내에서 명사 및 형용사와 같은 단어 경계와 레이블을 식별할 수 있음

**샘플 텍스트:** 빠르고, 새로운 저의 Kindle Fire가 좋아요!

[출처] <https://aws.amazon.com>

텍스트	태그
I	대명사
좋다	동사
저의	대명사
신속함	형용사
,	구두점
신규	형용사
Kindle	대명사
Fire	대명사
!	구두점

- 언어 감지
  - ✓ 100개 이상의 언어로 작성된 텍스트를 자동으로 식별하고, 신뢰도 점수와 함께 감지된 언어를 반환

**샘플 텍스트:** Amazon Elastic Compute Cloud (Amazon EC2) è un servizio Web che fornisce capacità di elaborazione sicura e scalabile nel cloud. È concepito per rendere più semplice il cloud computing su scala Web per gli sviluppatori.

ISO-639-1 언어 코드	언어	신뢰도
it	이탈리아어	1.0

[출처] <https://aws.amazon.com>

- 고객 분류

- ✓ ML을 배우지 않고도 사업 고유의 레이블로 사용자 지정 텍스트분류 모델을 손쉽게 빌드할 수 있음

**예:** 어떤 항공사에서 고객 지원 피드백을 정리하려 한다고 가정하겠습니다. 각각의 피드백을 계정 관련 질문, 항공권 환불, 비행 불만으로 분류하고 싶습니다. 이 서비스를 훈련 시키기 위해 사안별 예제 텍스트가 들어 있는 CSV 파일을 만들고, 각각의 샘플에 해당하는 네 가지 레이블 중 하나를 지정합니다. 그러면 사용자 대신 이 서비스가 사용자 지정 모델을 자동으로 훈련시킵니다. 다음 날, 이 모델을 사용하여 모든 통화를 분석하기 위해 각 텍스트 파일을 서비스에 제출했더니 레이블 지정된 결과와 함께 레이블 일치 신뢰도가 돌아왔습니다.

[출처] <https://aws.amazon.com>

텍스트	레이블	신뢰도 점수
행0	계정 관련 질문	0.92
행1	항공권 환불	1
행2	비행 불만	1
행3	비행 불만	0.91
Doc5.csv	항공권 환불	1

- 엔터티 인식

- ✓ 입력된 텍스트를 기반으로 자동 분류된 명명된 엔터티("사람", "장소", "위치" 등)를 반환

**샘플 텍스트:** 1994년 7월 5일 제프 베조스가 설립한 Amazon.com, Inc.는 워싱턴주 시애틀에 있으며, 책에서 블렌더까지 모든 품목을 판매합니다. 시애틀은 포틀랜드의 북쪽, BC 밴쿠버의 남쪽에 있습니다. 시애틀에 위치한 다른 회사로는 Starbucks와 Boeing이 있습니다.

[출처] <https://aws.amazon.com>

개체	범주	개수	신뢰도
Amazon.com, Inc.	Organization	1	0.96
워싱턴 시애틀	위치	1	0.96
1994년 7월 5일	날짜	1	0.99
제프 베조스	사람	1	0.99
Seattle	위치	2	0.98
포틀랜드	위치	1	0.99
브리티시 컬럼비아주 밴쿠버	위치	1	0.97
Starbucks	Organization	1	0.91
Boeing	Organization	1	0.99

- 언어 감지

- ✓ 사용자 지정 엔터티로 Amazon Comprehend를 사용자 지정하여 해당 도메인 특유의 용어를 찾아볼 수 있음

**샘플 텍스트:** 안녕하세요, 저는 Sam Ford라고 하는데 자동차 사고 보험금 청구서를 작성 중입니다. 제 정책 코드는 456-YQT이고요.

엔터티	범주	개수	신뢰도
456-YQT	Policy_ID	1	0.95

[출처] <https://aws.amazon.com>

- 다중 언어 지원

- ✓ 영어, 프랑스어, 독일어, 이탈리아어, 포르투갈어 및 스페인어 텍스트에 대한 텍스트 분석을 수행
- ✓ 따라서 여러 언어로 된 텍스트를 인식하고 Amazon Translate를 통해 텍스트를 영어, 프랑스어, 독일어, 이탈리아어, 포르투갈어 및 스페인어로 변환한 다음 Amazon Comprehend를 사용해 텍스트를 분석할 수 있는 애플리케이션을 만들 수 있음



- 주제 모델링

- ✓ Amazon S3에 저장된 문서 집합에서 관련 용어 또는 주제를 식별

**예시:** 문서(Doc1.txt, Doc2.txt, Doc3.txt 및 Doc4.txt)가 Amazon S3에 저장되어 있고 Amazon Comprehend에 문서의 위치를 지시하면 Comprehend가 문서를 분석하고 다음 두 가지 보기를 반환합니다.

#### 1. 주제 키워드의 그룹화.

각 키워드 그룹은 주제 그룹과 연결됩니다. 가중치는 그룹 내에서 해당 키워드의 사용 빈도를 나타냅니다. 키워드의 가중치가 10에 가까울수록 주제 그룹의 컨텍스트를 가장 잘 나타냅니다.

주제 그룹	키워드	무게
1	Amazon	0.87
1	Seattle	0.65
2	공휴일	0.78
2	쇼핑	0.67

#### 2. 주제별 문서 그룹화.

문서 이름	주제 그룹	비율
Doc1.txt	1	0.87
Doc2.txt	1	0.65
Doc3.txt	2	0.78
Doc4.txt	2	0.67



실습. Comprehend를 이용하여 텍스트에서 감정 분석

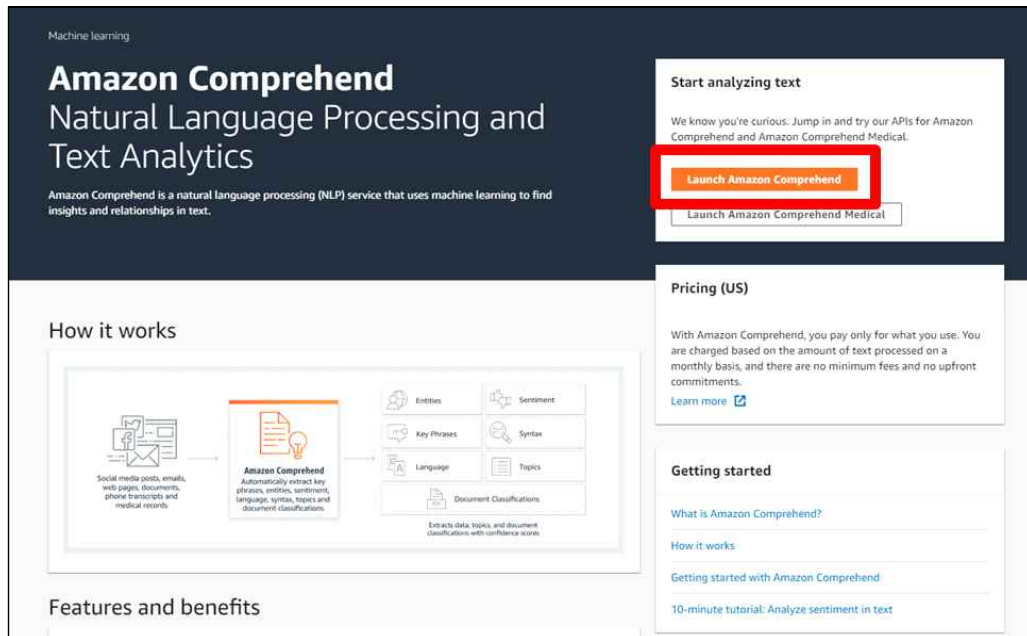
- Amazon Comprehend 시작하기
  - ✓ Amazon Comprehend의 감정 분석 기능을 둘러보면서 서적에 대한 후기 2개의 감성을 파악하고 책 구매 의사를 결정
  - 후기 1

I just wanted to find some really cool new places I've never visited before but no luck here. Some of these suggestions are just terrible... I had to laugh! Most Suggestions were just your typical big cities, restaurants and bars. Nothing off the beaten path here. I don't want to go these places for fun. Totally not worth getting this.

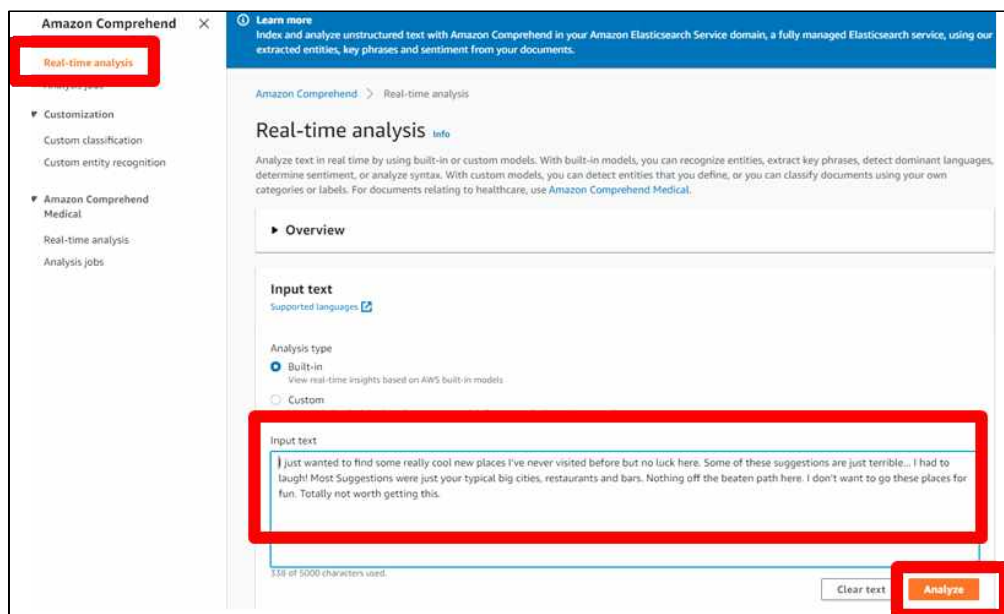
- 후기 2

This was such a beautiful book. I wasn't even planning and travel when I came across this and just started flipping through the pages. I really like the cover and all the large glossy photographs in this book. John Smith did a wonderful job with the photography. I've found a perfect home for this on my coffee table. I'm planning a trip to Paris and Barcelona soon and I know this will come in handy. In the meantime, it's perfect for assisting this armchair traveler!

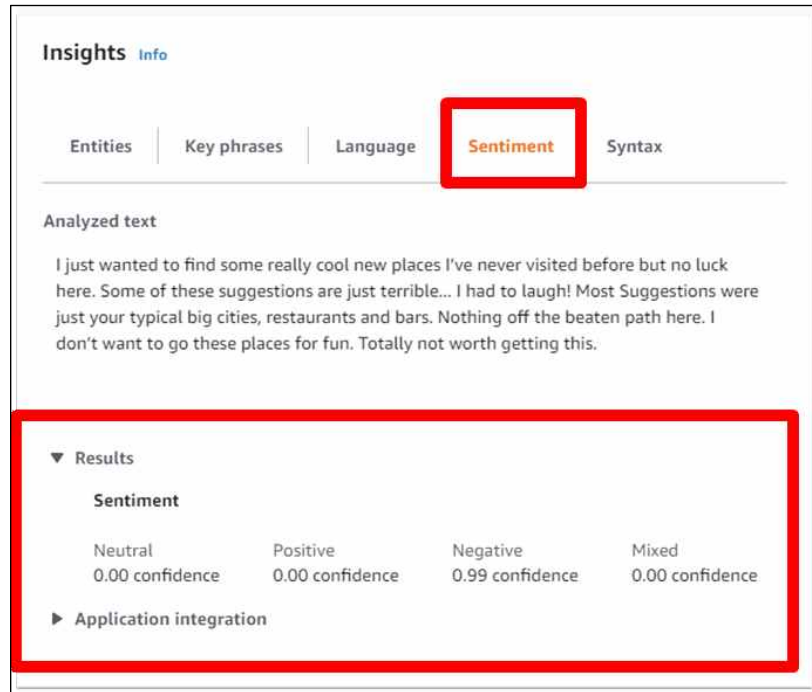
- ✓ Amazon Comprehend 콘솔에 접속하여 Launch Amazon Comprehend 시작



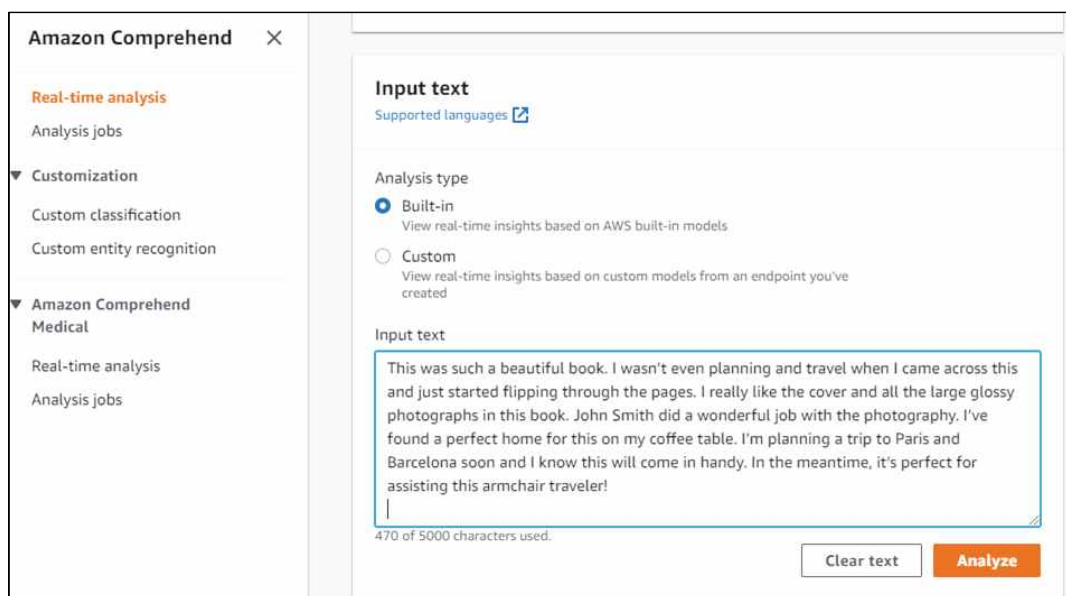
- 후기 1의 텍스트 입력 및 분석
  - ✓ 고객 후기가 긍정적인지, 부정적인지, 혼합적인지 분석하기
    - 왼쪽 대시보드에서 [Real-time analysis] 클릭 후 input text에 후기 1 입력 후 [분석] 클릭



- ✓ Insight 에서 Sentiment 탭을 선택하여 감정 확인
  - 후기가 부정적인 것을 확인할 수 있음



- 후기 2의 텍스트 입력 및 분석
  - ✓ 고객 후기가 긍정적인지, 부정적인지, 혼합적인지 분석하기
    - 왼쪽 대시보드에서 [Real-time analysis] 클릭 후 input text에 후기 2 입력 후 [분석] 클릭



## ✓ 감정 분석 사이드 바 열기

**Insights** Info

Entities | Key phrases | Language | **Sentiment** | Syntax

---

**Analyzed text**

This was such a beautiful book. I wasn't even planning and travel when I came across this and just started flipping through the pages. I really like the cover and all the large glossy photographs in this book. John Smith did a wonderful job with the photography. I've found a perfect home for this on my coffee table. I'm planning a trip to Paris and Barcelona soon and I know this will come in handy. In the meantime, it's perfect for assisting this armchair traveler!

▼ **Results**

**Sentiment**

Neutral	Positive	Negative	Mixed
0.00 confidence	0.99 confidence	0.00 confidence	0.00 confidence

▶ Application integration

## ✓ 엔터티 탐지 사이드바 열기

**Entities** | Key phrases | Language | Sentiment | Syntax

---

**Analyzed text**

This was such a beautiful book. I wasn't even planning and travel when I came across this and just started flipping through the pages. I really like the cover and all the large glossy photographs in this book. John Smith did a wonderful job with the photography. I've found a perfect home for this on my coffee table. I'm planning a trip to Paris and Barcelona soon and I know this will come in handy. In the meantime, it's perfect for assisting this armchair traveler!

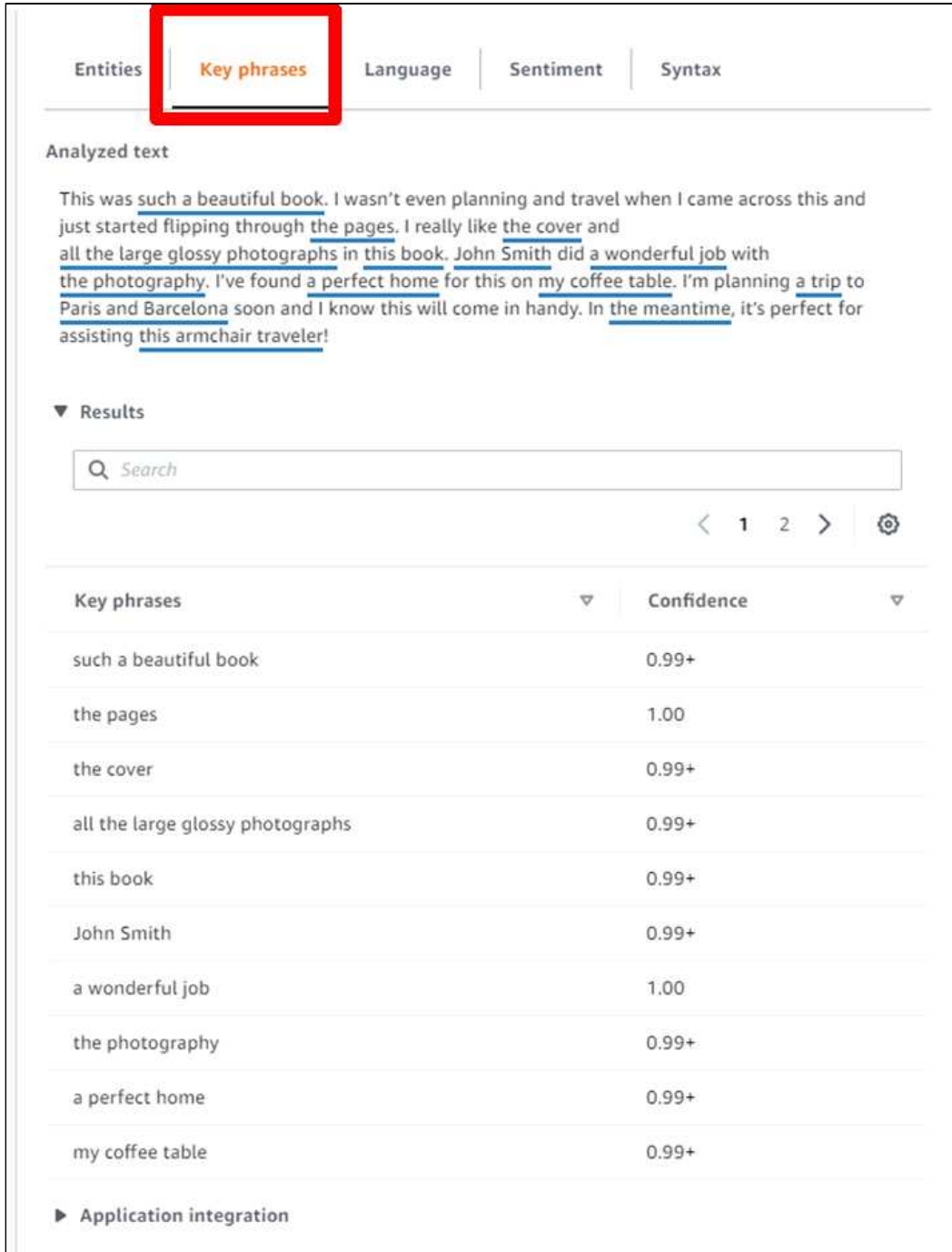
▼ **Results**

< 1 > ⚙

Entity	Category	Confidence
John Smith	Person	0.99+
Paris	Location	0.99+
Barcelona	Location	0.99+

▶ Application integration

✓ 핵심 문구 추출 사이드 바 열기



The interface shows a text analysis tool with tabs for Entities, Key phrases, Language, Sentiment, and Syntax. The 'Key phrases' tab is selected. Below the analyzed text, a search bar and navigation controls are present. A table displays the extracted key phrases and their confidence scores.

Key phrases	Confidence
such a beautiful book	0.99+
the pages	1.00
the cover	0.99+
all the large glossy photographs	0.99+
this book	0.99+
John Smith	0.99+
a wonderful job	1.00
the photography	0.99+
a perfect home	0.99+
my coffee table	0.99+

Application integration

**평가하기**

1. Amazon Comprehend의 기능이 아닌 것은?

- ① 핵심 문구 추출
- ② 구문 분석
- ③ 언어 감지
- ④ 이미지 분류

- 정답 : ④ 분

해설 : Amazon Comprehend의 기능으로는 핵심 문구 추출, 감성 분석, 구문 분석, 언어 감지, Custom Classification, 엔터티 인식, 사용자 지정 엔터티, 다중 언어 지원, 주제 모델링이 있습니다.

2. 토큰화 및 품사(PoS)를 사용하여 텍스트를 분석하고 명사 및 형용사와 같은 단어 경계를 분석하는 Comprehend 기능은?

- ① 구문 분석
- ② 감성 분석
- ③ 엔터티 인식
- ④ 언어 감지

- 정답 : ① 번

해설 : 구문 분석은 토큰화 및 품사(PoS)를 사용하여 텍스트를 분석하고 텍스트 내에서 명사 및 형용사와 같은 단어 경계와 레이블을 식별할 수 있습니다.

## 학습정리

### 1. Amazon Comprehend 개요

- 머신러닝을 사용하여 텍스트 안에 있는 통찰력과 관계를 찾아내는 자연어 처리 서비스
- 장점
  - 텍스트에서 더 나은 답을 확보
  - 주제별 문서 구성
  - 자체 데이터에 대한 훈련 모델
  - 일반 및 특정 텍스트 지원

### 2. Amazon Comprehend 주요기능

- 핵심 문구 추출
- Custom Classification
- 감성 분석
- 엔터티 인식
- 구문 분석
- 사용자 지정 엔터티
- 언어 감지
- 다중 언어 지원
- 주제 모델링

### 3. 실습. Comprehend를 이용하여 텍스트에서 감성 분석

- Amazon Comprehend 시작하기
- 후기 1의 텍스트 입력 및 분석
- 후기 2의 텍스트 입력 및 분석