



# Azure에서 Text Analytics 시작하기

3분

Text Analytics 서비스는 원시 텍스트를 통해 고급 자연어 처리를 수행할 수 있는 Azure Cognitive Services 제공 서비스의 일부입니다.

## Text Analytics 서비스용 Azure 리소스

애플리케이션에서 Text Analytics 서비스를 사용하려면 Azure 구독에 적절한 리소스를 프로비저닝해야 합니다. 다음 종류의 리소스 중 하나를 프로비저닝하도록 선택할 수 있습니다.

- **Text Analytics 리소스** - Text Analytics 서비스를 사용하려는 경우 또는 다른 서비스와 별도로 리소스에 대한 액세스 및 청구를 관리하려는 경우, 이 리소스 유형을 선택합니다.
- **Cognitive Services 리소스** - 다른 인지 서비스와 함께 Text Analytics 서비스를 사용하려는 경우 이 리소스 종류를 선택하고, 이 서비스에 대한 액세스 및 청구를 함께 관리하려고 합니다.

## 언어 감지

Text Analytics 서비스의 언어 감지 기능을 사용하여 텍스트가 작성된 언어를 식별합니다. 분석을 위해 한 번에 여러 문서를 제출할 수 있습니다. 제출된 각 문서에 대해 서비스는 다음을 검색합니다.

- 언어 이름(예: "영어")
- ISO 6391 언어 코드(예: "en")
- 언어 감지에 대한 신뢰 수준을 나타내는 점수

예를 들어 고객이 설문조사를 완료하고 음식, 서비스, 직원 등에 대한 피드백을 제공할 수 있는 레스토랑을 소유하고 운영하는 시나리오를 생각해 보겠습니다. 고객으로부터 다음과 같은 리뷰를 받았다고 가정해 보세요.

**리뷰 1 :** " A fantastic place for lunch. The soup was delicious. "

**리뷰 2 :** " Comida maravillosa y gran servicio. "

**리뷰 3 :** " The croque monsieur avec frites was terrific. Bon appetit! "

Text Analytics 서비스를 사용하여 이러한 각 리뷰에 대한 언어를 검색할 수 있고, 서비스에서 다음과 같은 결과로 응답할 수 있습니다.

문서	언어 이름	ISO 6391 코드	점수
리뷰 1	영어	en	1.0
리뷰 2	스페인어	es	1.0
리뷰 3	영어	en	0.9

영어와 프랑스어가 혼합된 텍스트이지만 리뷰 3에서 감지된 언어는 영어입니다. 언어 감지 서비스는 텍스트에서 **지배적** 언어에 중점을 둡니다.\*\_ 이 서비스는 알고리즘을 사용하여 텍스트의 다른 언어와 비교하고 언어에 대한 구 길이 또는 총 텍스트 양과 같은 지배적 언어를 결정합니다. 지배적 언어는 언어 코드와 함께 값이 반환됩니다. 혼합 언어 텍스트의 결과로 신뢰도 점수가 1 미만일 수 있습니다.

## 모호하거나 혼합된 언어 콘텐츠

속성이 모호하거나 언어 콘텐츠가 혼합된 텍스트가 있을 수 있습니다. 이러한 상황은 서비스에 문제를 일으킬 수 있습니다. 모호한 콘텐츠 예시는 문서에 제한된 텍스트 또는 문장 부호만 포함하는 경우입니다. 예를 들어 서비스를 사용하여 “:-)”라는 텍스트를 분석하면 언어 이름과 언어 식별자 값이 \_알 수 없음 \*이고, 점수는 ‘숫자가 아님’을 뜻하는 NaN 으로 표시됩니다.

## 감정 분석

Text Analytics 서비스는 텍스트를 평가하고 각 문장에 대한 감정 점수 및 레이블을 반환할 수 있습니다. 이 기능은 소셜 미디어, 고객 리뷰, 토론 포럼 등에서 긍정적이고 부정적인 감정을 감지하는 데 유용합니다.

이 서비스는 미리 빌드된 기계 학습 분류 모델을 사용하여 텍스트를 평가하고 0에서 1 범위의 감정 점수를 반환하는데, 1에 가까운 값일수록 긍정적인 감정을 나타냅니다. 범위의 중간(0.5)에 가까운 점수는 중립 또는 불확실로 간주합니다.

예를 들어, 다음 레스토랑 리뷰 두 개의 감정을 분석할 수 있습니다.

“어젯밤 이 레스토랑에서 저녁 식사를 했으며 직원들의 정중한 태도가 첫눈에 들어왔습니다. 직원들은 우리에게 친절한 태도로 인사를 하고 테이블로 곧장 안내했습니다. 테이블은 깨끗했고, 의자는 편안했으며, 음식은 정말 맛있었습니다.”

그리고

“이 레스토랑에서 경험한 저녁 식사는 최악 중 하나였습니다. 서비스가 느렸고 음식은 형편 없었습니다. 다시는 여기에서 식사하지 않을 것입니다.”

첫 번째 리뷰에 대한 감정 점수는 약 0.9일 수 있으며, 이는 긍정적인 감정을 나타냅니다. 한편 두 번째 리뷰 점수는 0.1에 가까울 수 있으며, 이는 부정적인 감정을 나타냅니다.

## 불확실한 감정

점수 0.5는 텍스트의 감정이 불확실함을 나타낼 수 있으며, 이는 감정을 드러낼 만한 충분한 맥락이 없거나 불충분한 구문의 텍스트로 인해 발생할 수 있습니다. 예를 들어 구조가 없는 문장의 단어 목록은 불확실한 점수를 유발할 수 있습니다. 점수가 0.5일 수 있는 또 다른 예는 잘못된 언어 코드를 사용한 경우입니다. 언어 코드(예: 영어의 경우 "en", 프랑스어의 경우 "fr")는 텍스트에 사용된 언어를 분석 서비스에 알리는 데 사용됩니다. 프랑스어로 텍스트를 전달하지만 서비스에 언어 코드를 영어인 **en** 이라고 알려준다면 서비스는 정확하게 0.5의 점수를 반환합니다.

## 핵심 구 추출

핵심 구 추출은 문서 한 개 또는 여러 개의 텍스트를 평가한 다음 문서의 주요 논점을 파악하는 개념입니다. 앞서 설명한 레스토랑 시나리오를 생각해보겠습니다. 수집한 설문조사의 양에 따라 리뷰를 전부 읽는 데 시간이 오래 걸릴 수 있습니다. 그 대신 Text Analytics 서비스의 핵심 구 추출 기능을 사용해서 주요 논점을 요약할 수 있습니다.

다음과 같은 리뷰를 받을 수 있습니다.

“여기서 생일 축하를 위해 저녁 식사를 했고 환상적인 경험이었습니다. 친절한 여주인이 우리에게 인사를 하고 바로 테이블로 데려갔습니다. 분위기는 편안했고, 음식은 훌륭했으며, 서비스 또한 뛰어났습니다. 훌륭한 음식과 세심한 서비스를 원한다면 이 장소를 추천합니다.”

핵심 구 추출은 다음과 같은 구를 추출하여 이 리뷰에 몇 가지 컨텍스트를 제공할 수 있습니다.

- 세심한 서비스
- 훌륭한 음식
- 생일 축하
- 환상적인 경험
- 테이블
- 친절한 여주인
- 저녁 식사
- 분위기
- 장소

감정 분석을 사용하여 이 리뷰가 긍정적이라고 판단할 수 있을 뿐만 아니라, 핵심 구를 사용하여 리뷰의 중요한 요소를 식별할 수 있습니다.

# 엔터티 인식

Text Analytics 서비스에 구조화되지 않은 텍스트를 제공하면, 인식한 텍스트의 '엔터티' 목록을 반환합니다. 이 서비스는 웹상의 해당 엔터티에 대한 보다 자세한 정보로 연결하는 링크를 제공할 수도 있습니다. 엔터티는 기본적으로 특정 유형 또는 범주의 항목입니다. 경우에 따라 다음 표에 나온 것과 같은 하위 유형이 있습니다.

유형	하위 유형	예
사람		"빌 게이츠", "존"
위치		"파리", "뉴욕"
조직		"Microsoft"
수량	숫자	"6" 또는 "여섯"
수량	백분율	"25%" 또는 "오십 퍼센트"
수량	서수	"1st" 또는 "첫 번째"
수량	나이	"생후 90일" 또는 "30세"
수량	통화	"10.99"
수량	차원	"10마일", "40cm"
수량	온도	"45도"
DateTime		"2012년 2월 4일, 오후 6:30"
DateTime	날짜	"2017년 5월 2일" 또는 "2017/05/02"
DateTime	시간	"8am" 또는 "8:00"
DateTime	DateRange	"5월 2일~5월 5일"
DateTime	TimeRange	"오후 6시부터 오후 7시"
DateTime	기간	"1분 45초"
DateTime	설정	"매주 화요일"
URL		"https://www.bing.com"

유형	하위 유형	예
메일		"support@microsoft.com"
미국 국내 전화 번호		"(312) 555-0176"
IP 주소		"10.0.1.125"

또한 이 서비스는 특정 참조에 연결하여 엔터티를 구분하도록 도와주는 '엔터티 링크' 설정을 지원합니다. 인식된 엔터티에 대해 서비스는 관련 'Wikipedia' 아티클 URL을 반환합니다.

예를 들어, Text Analytics 서비스를 사용하여 다음 레스토랑 리뷰 추출의 엔터티를 검색한다고 가정해 보겠습니다.

"지난주에 시애틀에 있는 레스토랑에서 식사했습니다."

엔터티	유형	하위 유형	Wikipedia URL
시애틀	위치		<a href="https://en.wikipedia.org/wiki/Seattle">https://en.wikipedia.org/wiki/Seattle</a>
지난주	DateTime	DateRange	

## 다음 단원: 연습 - Text Analytics 서비스를 사용한 텍스트 분석

계속 >