

# Cómo hacer un keyword research con **nlp** y **spacy**

Técnicas de extracción de entidades para el SEO moderno.

Basado en los datos de David Martínez Mata (Los Creativos).

# El problema del keyword research tradicional

Caso de estudio: Consulta 'aprender inglés' en Colombia (Ahrefs).

12.800 términos asociados encontrados.

**94%** (12.024 keywords) marcan un volumen de búsqueda de cero.

66% del volumen total se agrupa en solo las 100 primeras posiciones.

| Keyword                           | Intents | KD | SV   | GSV  |
|-----------------------------------|---------|----|------|------|
| + aprender ingles                 | I C     | 0  | 3.1K | 34K  |
| + como aprender ingles            | I C     | 31 | 1.6K | 117K |
| + aprender ingles gratis          | I C     | 62 | 1.5K | 11K  |
| + como aprender ingles mas rapido | I C     | 4  | 700  | 1.2K |
| + aprender ingles                 | I       | 0  | 700  | 4.0K |
|                                   | I       | 9  | 500  | 15K  |
|                                   | I C     | 28 | 500  | 3.8K |
|                                   | I C     | 31 | 450  | 4.7K |
|                                   | I C     | 0  | 450  | 6.0K |
|                                   | I       | 0  | 400  | 3.3K |
|                                   | I       | 48 | 400  | 3.9K |
|                                   | I       | 0  | 350  | 3.0K |
|                                   | I C     | 44 | 350  | 3.1K |
|                                   | I       | 51 | 250  | 1.1K |
|                                   | I C     | 31 | 250  | 2.1K |
|                                   | I C     | 33 | 250  | 2.4K |
|                                   |         |    |      | 3.0K |
|                                   |         |    |      | 2.9K |
|                                   |         |    |      | 2.5K |

Los modificadores dominantes (gratis, rápido, fácil, app) entregan promesas transaccionales, no conocimiento estructurado.

# La evolución: de cadenas de texto a significados

| Época     | Modelo                 | Límite / SEO                                                                            |
|-----------|------------------------|-----------------------------------------------------------------------------------------|
| 1960-1990 | Booleano / Vectorial   | Coincidencia exacta. Si no estaba la palabra, no existías. (SEO: Densidad de keywords). |
| 1998      | PageRank               | Autoridad vía enlaces. (SEO: Links).                                                    |
| 2012      | Knowledge Graph        | <b>Conecta entidades</b> . Things, not strings. (SEO: <b>Marcado semántico</b> ).       |
| 2015-2021 | BERT, MUM (Neuronales) | Búsqueda de intención y significado.                                                    |
| 2022+     | LLMs                   | Reconstrucción estadística. Representación basada en probabilidad y autoridad.          |

En 2012, Google dejó de buscar coincidencias exactas para buscar cosas (entidades).

# Keywords vs. entidades: el cambio de enfoque

## La Consulta

**Concepto:** Keyword / Cadena de texto.

**Naturaleza:** Lo que el usuario escribe. Mide la demanda (ej. "aprender inglés gratis").

**Problema:** Se multiplica en infinitas variaciones superficiales.

aprender inglés  
como aprender inglés gratis  
curso de inglés rápido  
aprende ingles ahora  
mejor app para ingles  
quiero hablar ingles  
ingles desde cero  
donde estudiar ingles  
clases de inglés  
clases de inglés

## El Referente

**Concepto:** Entidad / NLP.

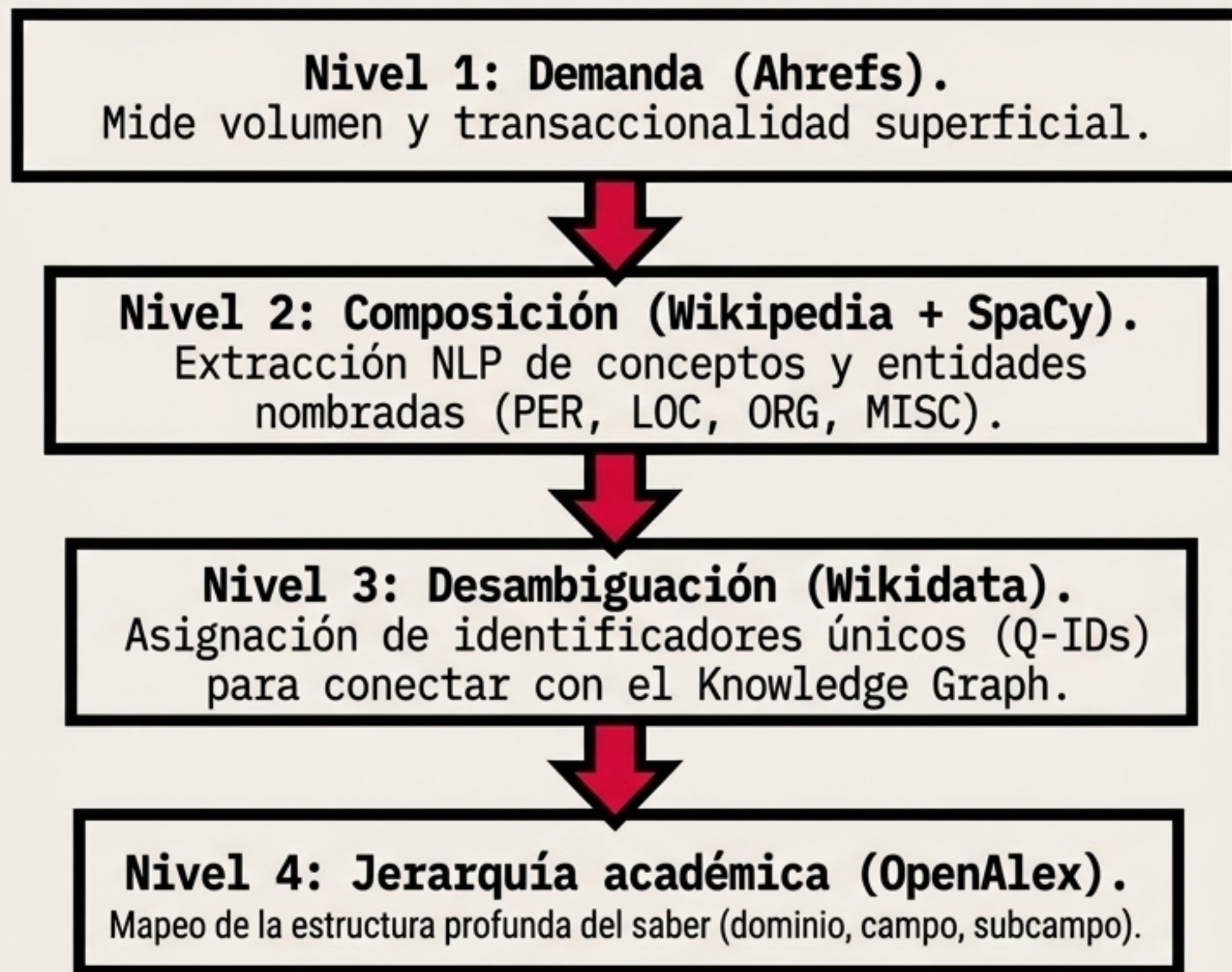
**Naturaleza:** Una "cosa" única en el mundo (ej. "Idioma inglés", "MCER", "TOEFL"). Mapea el significado.

**Ventaja:** Un LLM no busca el término exacto; construye la respuesta evaluando relaciones, contexto y autoridad.



**"Aprender inglés" no es una entidad. Pero está hecha de entidades.**

# Metodología de extracción en 4 niveles



# De la transacción a la composición del saber

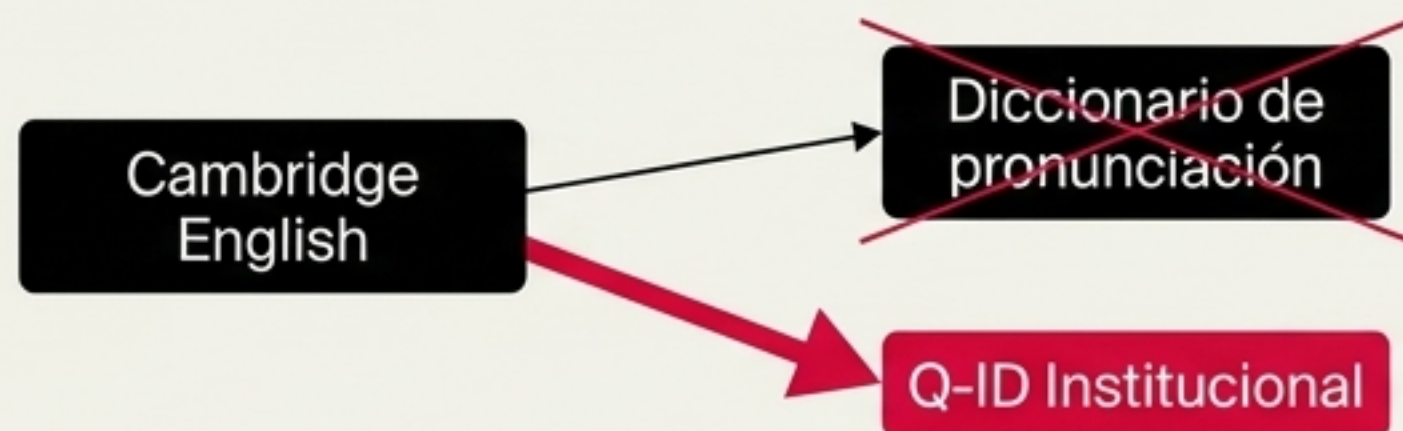
|                                                                                               |
|-----------------------------------------------------------------------------------------------|
| <b>Nivel 1 - Demanda (Ahrefs)</b>                                                             |
| Modificadores dominantes en la búsqueda: gratis <b>(1.417)</b>   rápido   fácil   app   niños |

|                                                    |                                            |                                             |
|----------------------------------------------------|--------------------------------------------|---------------------------------------------|
| <b>Nivel 2 - Extracción (SpaCy)</b>                |                                            |                                             |
| Lo que subyace al tema al analizar textos con NLP: |                                            |                                             |
| <b>Conceptos (Sintagmas)</b>                       | <b>ORG (Organizaciones)</b>                | <b>MISC (Miscelánea)</b>                    |
| Bilingüismo, hablantes nativos, lengua inglesa.    | British Council, Universidad de Cambridge. | IELTS, Duolingo, Oxford English Dictionary. |

**SpaCy revela quién certifica y dónde ocurre, desplazando el foco de la transacción a la autoridad.**

# Conexión con el knowledge graph y la academia

## Nivel 3 - Wikidata / Q-IDs



El nombre no basta: la IA apunta a diccionarios por error.

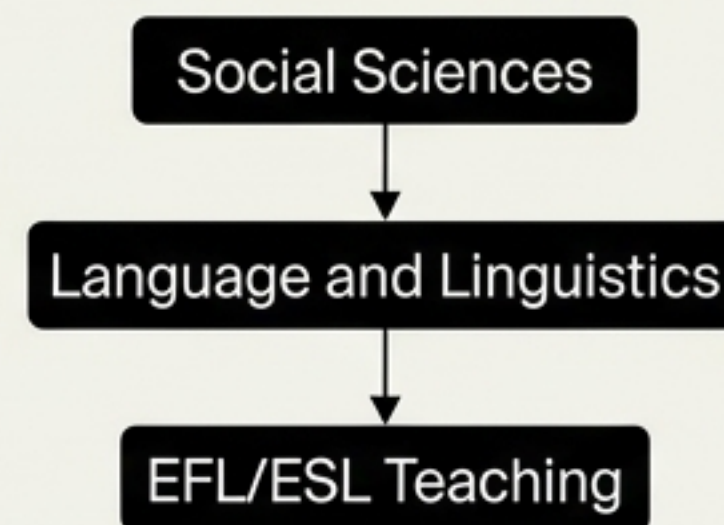
Solución: Desambiguar con Q-IDs exactos.

**Q2918660** (Duolingo)

**Q487425** (TOEFL)

**Q221385** (MCER)

## Nivel 4 - OpenAlex / Jerarquía



Análisis de 125 trabajos académicos revela la estructura profunda.

Conceptos revelados:

- **Hipótesis del periodo crítico**
- **Evaluación de competencia**
- **Función ejecutiva**

# El volumen de cero no significa que nadie pregunte

| Concepto académico (Profundidad) | Pregunta real de usuario (Ahrefs)                          |
|----------------------------------|------------------------------------------------------------|
| Critical period hypothesis       | ¿A qué edad se puede aprender inglés? <b>(Vol. 0)</b>      |
| Language immersion               | ¿Cómo aprender inglés como un nativo? <b>(Vol. 0)</b>      |
| Language proficiency assessment  | ¿Cuántos niveles hay para aprender inglés? <b>(Vol. 0)</b> |
| Adult second language learning   | ¿Cómo aprender inglés siendo adulto? <b>(Vol. 0)</b>       |

Las herramientas son ciegas a conversiones LLM. Si buscas autoridad, una pregunta de volumen cero sobre un concepto central vale más que diez variaciones de la palabra "gratis".

# Cómo cambia la estructura de un artículo

Contexto: Arquitectura para el artículo "¿A qué edad se puede aprender inglés?"

## Antes (Basado en volumen superficial)

H2: Aprender rápido siendo adulto

H2: Apps gratis para aprender

H2: Tips fáciles desde casa

*(Diagnóstico: Repetición de la cadena de texto, sin respuesta real).*

## Después (Basado en entidades y OpenAlex)

H2: ¿Existe una edad límite?

— [Entidad: Hipótesis del periodo crítico]

H2: ¿Qué nivel necesito?

└ [Entidad: Q221385 - MCER]

└ H3: ¿Qué examen certifica mi nivel?

└ [Entidad: Q487425 - TOEFL]

H2: Beneficios cognitivos

└ [Entidad: Bilingüismo y función ejecutiva]

# 10 pasos para pasar de palabras a entidades

