<https://rapidapi.com/UnFound/api/keywords-extraction2> - Api para buscar palabras claves en textos. Podría servir para recomendar etiquetas. Nota: La versión gratuita admite solo hasta 100 requests, y es necesario meter tarjeta.

<https://www.programmableweb.com/category/keywords/api> - Varias api para buscar palabras clave

<https://www.quora.com/What-are-the-most-popular-text-similarity-algorithms> - Respuesta sobre algoritmos para buscar que tan similares son dos textos (para buscar pregunta repetida / relacionada)

<https://es.wikipedia.org/wiki/Tf-idf> Algoritmo de similitud de textos

<https://meta.stackexchange.com/questions/20473/how-are-related-questions-selected> - Como maneja Stack Overflow las preguntas relacionadas:

* En 2009: Essentially, it's a vocabulary search - which other posts have the same vocabulary as this one.

Of course, there's a lot of tuning involved - the top 10k most common English words are excluded, and tags are included, plus (probably) some magic sauce they haven't revealed.

As far as I've heard, there's nothing you can do if a a question on the list is really unrelated, as the matching is entirely automated, done in real time (no preprocessing, means nothing to tweak).

* Actualmente: Things have changed a bit since 2009 :-) We rely heavily on Elasticsearch these days. The "related" questions now come from an [Elasticsearch "More Like This" query](https://www.elastic.co/guide/en/elasticsearch/reference/current/query-dsl-mlt-query.html).

Elastic uses [tf-idf](https://en.wikipedia.org/wiki/Tf%E2%80%93idf) for most things, which is a standard search algorithm. Their particular implementation is [documented here](https://lucene.apache.org/core/4_9_0/core/org/apache/lucene/search/similarities/TFIDFSimilarity.html).

<https://towardsdatascience.com/extracting-keywords-from-short-text-fce39157166b> - Extraer palabras clave de un texto (cosas algo avanzadas)

Podría ser que cada pregunta tuviera una lista de (hasta 5?) palabras clave. Estas se obtendrían buscando las palabras que mas se repiten en la pregunta menos las palabras que claramente se repiten (Como conjunciones y cosas así. estas se podrían encontrar buscando cuales son las palabras que mas se repiten del idioma)