White Paper – How Kvasir's Vector Mapping Technology is Positioned Against Other Search Technologies

1. Introduction/Overview

Kvasir uses statistical vector mapping techniques for knowledge management in contrast to most other knowledge management products available on the market today which predominantly base their services on rules-based NLP technology.

Kvasir's technology is therefore not appropriate for all enterprise knowledge management scenarios, but it has significant advantages in certain circumstances. This paper discusses the advantages and disadvantages of Kvasir's vector mapping technology against other methods and highlights the sectors and use cases where this approach can bring significant advantages to organisations.

2. Overview of competing technologies

2.1. Kvasir's Vector Mapping Technology

Kvasir's core Vector Mapping technology is based on techniques developed at Cambridge University's Department of Computer Science and Technology.

Vector Mapping ingests documents - where the definition of a document is just a collection of words ranging from tweets to PDFs to closed captioning on YouTube videos - before applying neural NLP techniques to convert each document into a 300 dimensional vector, and then creating indexes that enable efficient nearest-neighbours search of the resulting vector space. The combination of the set of vectors, any associated metadata, and the indexes of the resulting vector space is called a "knowledge pack".

Queries are formed by ingesting an input document in a similar way to represent it as a 300 dimensional vector and then using k-nearest-neighbours and Word Movers Distance to find and return to the user the closest-matching documents in the specified knowledge packs.

The advantage of this approach is that each ingested document is treated simply as a collection of words and there is no reliance on structure, linking, metadata, ontologies or taxonomies to index a document. This means very unstructured data can be indexed, brought into knowledge packs, and the total knowledge pool of an organisation extended. When querying, an entire document can then be used as a source for the query as opposed to just a few keywords which significantly increases match accuracy and relevance.

More detailed information on Kvasir's technology is available in a technical white paper which is available on request, and peer reviewed papers on the algorithms involved can be found at https://kvasira.com/technology).

2.2.   Natural Language Processing and keywords

Today, typical use of Natural Language Processing (NLP) in knowledge management focuses on rules-based NLP, enabling users to create queries based on questions in natural language, e.g., "what is the largest country in the world?".

By using an understanding of language structure (syntactic NLP) and combining this with semantic technology (symbolic NLP), this approach extracts keywords from the user's input to create a search query that can be fulfilled using standard keyword-based lookup techniques.

This approach works well for queries in the form of short questions or sentences but breaks down with larger and complex pieces of text where understanding sentence structure and extracting the most meaningful keywords becomes difficult.

2.3.   Knowledge Graphs

Knowledge graphs build relationships between documents to enable users to easily find information related to their search query. For example, if the search query was "William Shakespeare" this could be mapped to a node in a knowledge graph that might be linked to other nodes representing Shakespeare's plays, his wife Anne Hathaway, or his birthplace Stratford-upon-Avon.

The relationships in knowledge graphs are typically built by semantic and syntactic analyses of large text corpuses to extract information, e.g., the sentence "William Shakespeare was married to Anne Hathaway" would indicate a relationship between William Shakespeare and Anne Hathaway, allowing nodes (representing entities) and edges (representing relationships) to be added to the knowledge graph.

Information provided to a user through knowledge graphs really only works where the input query can be effectively mapped onto one or a very small number of nodes already in the knowledge graph. Large, complex input queries which cannot be accurately represented by the entities in the knowledge graph will not generate meaningful results.

2.4.   Collaborative Filtering and other search techniques

There are a number of other techniques that can be used to present information back to users, including:

- Personal history. Biasing search results based on what the individual looked for previously.
- Group history. Prioritising search results based on learnings from similar searches and what users actually clicked on.
- Personal data. Using user profiles to influence results, for example location ("restaurants near me"), and demographic data ranging from age, sex, political leaning as well as any other profile data that can be calculated from previous online activity and the online activity of friends.
- Merchandising. Prioritising search results to optimise revenue, for example prioritising goods that are being promoted on an eCommerce site, or maximising advertising revenues.

The above techniques work well in a mass-market scenario with millions of users but are less applicable in an enterprise environment where there isn't the breadth of data to create query profiles.

The use of personal history and profiles also has GDPR and privacy ramifications.

3. Advantages & Disadvantages of Vector Mapping

As is set out above, there are advantages and disadvantages to all knowledge management technologies and so it is common to use a combination of technologies to optimise results.

The advantages and disadvantages of Kvasir's technology are explored in more detail below.

3.1. Advantages

Indexes can be created from very unstructured data

Kvasir's Vector Mapping technology does not rely on structured taxonomies, ontologies, meta-data, or even correct grammar in the source documents in order to create an index. As long as data contains a collection of words this data can be indexed. This means that knowledge packs can be created not only from the information you would expect (files in SharePoint, Dropbox etc), but also from information such as chat sessions, videos (through closed captioning), and even recorded speech using third party speech to text conversion tools.

This greatly extends the enterprise's knowledge base and user's access to information.

Easy deployment and low maintenance cost for data indexes

Most knowledge management systems rely on some level of structure in the data sources in order to create indexes. This often requires the enterprise to spend time and effort in maintaining ontologies and ensuring the right metadata and tags are attached to input documents to identify information.

Although Kvasir's solutions will ingest metadata if it is available, they do not require it: Vector Mapping uses the document content itself to determine the best results. In fact, Kvasir's solutions can even help automatically create and curate metadata based on the document's content.

Not only does this save the enterprise time and cost in deploying a solution, Kvasir's technology is also not impacted by the degradation in quality of data sources which invariably happens in enterprises over time as users fail to adhere to standards or mis-file information.

More accurate and relevant search queries

By forming queries as entire documents, Kvasir's Vector Mapping technology can access much richer context around an input query when compared to a short sentence or handful of keywords as used in traditional knowledge management search tools.

Vector Mapping's use of entire documents or parts thereof as search queries allows Kvasir's AI technology to greatly increase search accuracy and relevance.

Proactive information provision

As Kvasir's technology uses fuller context for a query, it lends itself particularly well to anticipating the information needs of a user and then supplying this information dynamically to the user within their workflow.

For example, consider a user writing a commercial proposal or research paper. As the user is creating this document Kvasir's AI knowledge management solution can continually scan the content as it is being created and then proactively provide information relevant to what's being written, such as similar proposals or related research papers. This not only makes the user more efficient by providing very accurate information within their workflow without them having to stop to look for it, it also

reduces business risk since critical information may be presented that the user didn't even know existed and so would never have looked for.

Multi-language capability

Vector Mapping technology also has the ability to deal with different languages without requiring translation. Kvasir's technology identifies the language of the document being used and, using the neural networks for that language, maps it into an aligned vector space. Once complete, each document is represented simply as a collection of 300 numbers allowing any document in any language to be compared to any other.

This means a user can interrogate an archive created in one language with a query in another and still get accurate and relevant results.

3.2. Disadvantages

Natural language queries are not supported

Since Kvasir's Vector Mapping technology does not attempt to determine the underlying semantics of the query, querying an index with a question like "how fast is the fastest car?" will not generate meaningful results. Kvasir's technology is simply not designed to take in these types of queries.

Vector Mapping does support keyword searching, but if only a few keywords are used results will be no better than any other keyword search technology. With Kvasir's technology accuracy and relevancy of results increases the longer and more detailed the search query, whereas this decreases with other search technologies.

User profiles are not used

Vector mapping technology uses only the provided input data as the query. It does not use other user profile or demographic information to weight or promote results. This means input queries like "Indian restaurants near me" will not generate results since "near me" needs an understanding of geographical information which Kvasir's solution does not have. Likewise search results are not presented back based on learned user behaviour from users performing similar searches.

Not using personal profiles however means Kvasir's technology is GDPR compliant and not having results biased by previous personal or group experience can be useful especially if novel research is being conducted.

A new way of working

When referencing digital search people naturally envisage the Google search interface - a text box where a user can enter a few keywords or a simple question. Over 20 years of Google's Internet search dominance has trained people to search in this way. Although Kvasir's service supports keyword search by entering a few words into a text box, this does not make the best use of the technology, which increases search accuracy and relevancy as more data is used as a query, up to and including large multi-page documents.

For this reason, Kvasir's preferred method of exposing the technology is through tools embedded into the workflow which also has the advantage of showcasing Kvasir's ability to proactively show information by anticipating user needs.

4. Positioning of Kvasir's Solutions against other knowledge management technologies

4.1. Overall positioning

Kvasir's Vector Mapping technology has advantages which makes it a very powerful tool in some situations, but it also has some limitations which means it is not applicable to all environments.

Kvasir's solution is not designed to be a consumer facing tool, a competitor to Google, or an eCommerce tool where product merchandising is important. NLP tools combined with knowledge graphs and use of personal profile data remain the best solutions for this market. Neither is Kvasir's solution a replacement for generic enterprise search solutions where enterprise users expect to perform simple keyword/NLP searches.

Kvasir's technology is a *complementary* solution to these for data heavy organisations that rely on access to multiple, complex and disparate data sources such as research papers, regulations, contracts and product information in order to succeed. By integrating Kvasir's technology into user workflows, e.g., as plug-ins in Microsoft Word, Jira, or Slack, Kvasir's solutions can work in parallel with existing traditional enterprise search solutions to proactively provide information to the end user in real time based on the full context of what they are working on at that moment.

4.2. Example Use Cases

The Patent Attorney

Kvasir has already created knowledge packs from major patent archives world-wide. Kvasir can also create knowledge packs from patent attorneys' internal archives and create a knowledge pack.

As an invention disclosure comes in from a client, the attorney can simply drag & drop the disclosure document into Kvasir's tool to immediately find the most relevant existing patents and academic research, regardless of the language the disclosure was written and using on the entire content of the disclosure to find the most relevant results.

This can save the patent attorney a huge amount of time since they don't have to worry about which keywords and synonyms to use for their query, and neither do they have to translate these into other languages.

The Commercial Manager

A commercial manager in a large engineering consultancy receives a Request For Information (RFI) from a prospective client. As the commercial manager writes the response to the RFI in MS Word, Kvasir's technology can inspect the content being created and suggest useful information. This could include similar RFI responses that the company has already created which the commercial manager can simply reference or cut & paste from and so save a huge amount of time.

Suggested information can also include specific detailed information relating to the section that the manager is working on. For example, if the manager is writing the Health & Safety section, Kvasir's technology can reference information specifically related to that section. This means that if there is a new H&S regulation that the manager didn't know about, Kvasir's technology will bring that to the manager's attention so reducing business risk.

The Software Engineer

Kvasir's technology can help software engineers work on tasks or resolve problems faster and more efficiently by integrating Kvasir's technology into task and issue tracking systems such as Jira or Bugzilla.

For example, when an engineer goes to their task list in Jira which may contain the task plus any associated information and commentary, Kvasir's solution can look at the entire context of the task and present information that may help the engineer address the task from the many resources at hand. Multiple information sources could be referenced including similar tasks in the system that have already been resolved, internal group chats that may have been discussing similar tasks, or external knowledge sources such as Stack Overflow where this issue has already been resolved.

Interactively providing suggestions where the task they are addressing has already been resolved based on the full context of the issue at hand can save the engineer a huge amount of time and enable a better solution.

The Procurement Manager

As well as enabling enterprise search and proactively providing information to users, Kvasir's technology can allow users to dynamically and graphically explore complex data sets.

Consider a procurement manager in a large organisation; Kvasir can create knowledge packs detailing each of their suppliers and partners combining very disparate information such as contracts, tenders, and information from the organisation's web site. The manager can then navigate this information visually so they could, for example, map their supplier ecosystem against the organisation's wants and needs, discover new potential sources of supply, and uncover areas of under-supply and vulnerability.

5. Summary

Kvasir's Vector Mapping technology is not a replacement for current enterprise search technologies that rely on grammatical parsing of queries and knowledge graphs. Rather, it is a complementary solution that works alongside these technologies by providing information to end users within their workflow.

Kvasir's technology is aimed at data heavy enterprises that rely on rich, complex, and disparate data to perform their functions on a daily basis.

By increasing the size of the corporate knowledge base, increasing the accuracy and relevance of search results by using a fuller context of user interaction, and doing all this proactively within the workflow of the user, Kvasir's technology can save time, reduce business costs, and reduce risk by proactively providing critical business information that the user didn't even know they needed.

Advanced Knowledge Management for Enterprises:
- **Increase the size of your knowledge base**
- **Reduce time spent searching**
- **Save costs**
- **Reduce business risk**

Kvasir
KNOWLEDGE MANAGEMENT

**More information:
contact@kvasira.com**