



# A guide to choosing the right natural language processing solution

If you work in **healthcare, pharmaceuticals or biotechnology**, it's an uphill struggle to make sense of the mass of textual data that faces you daily: Journal articles, patents, lab notebooks, clinical reports, internal documents, health records—the list is endless.

Resource-intensive and potentially error-prone, traditional keyword research and manual scanning no longer represent a practical solution to the problem of finding and analyzing information.

Natural language processing (NLP) is an artificial intelligence (AI) technology that is proven to streamline the analysis of information. However, with a growing number of NLP tools available, how do you decide which is the right solution for your business?

This guide will help you understand the key capabilities to look for when choosing your NLP solution and vendor.

Here are the top 10 capabilities to look for:

1

## **Contextual analysis and advanced linguistics**

A good NLP solution should be able to recognize linguistic entities and extract relationships, and use semantic software to understand and detect the mention of a concept, no matter how it's expressed in the text—using relevant ontologies to ensure proper understanding of clinical, medical and scientific text.

Ideally the solution will allow document region or section searching, and be able to process the important data often contained in complex tables. You also want some level of normalization of the output so you can easily group and visualize data sets, load the results into data warehouses or data lakes, or use them to drive machine learning (ML) models.

2

## **Domain knowledge, medical and scientific understanding**

You want to choose a solution and NLP vendor that speaks your language. Experienced professional services and customer support teams with deep industry domain knowledge are essential to ensure maximum efficiency and productivity with an NLP solution in your environment.

A collaborative community forum can also foster sharing of ideas, strategies, queries and best practices in a non-competitive setting.

3

## **Interoperability**

It is important that any NLP vendor or solution you choose has an open architecture, so that adding and swapping components and integrating tools into enterprise workflows is easy. A RESTful Web Services API can support integration with document processing workflows, and an open search language supporting all NLP functionality will simplify the creation of extraction strategies. The system should allow integration of unstructured data with Master Data Management, data warehousing and analytics tools.

**4****Scalability and performance**

Any NLP system you implement needs to have the scalability and performance to handle current and future data loads and volumes. You may need to run the platform over tens of millions of documents, handling thousand-page documents and managing terminologies/ontologies with millions of terms—sometimes in real time.

The system should run on various architectures, whether standard, multicore, cluster or cloud, and work with data stores like Hadoop, Documentum and SharePoint. It should also provide a connector to run the system in a service-oriented environment to handle unpredictable and variable workflows, such as Extract, Transform, and Load (ETL), semantic enrichment and signal detection/alerting.

**5****Deployment options**

You will need to consider how the NLP solution is deployed, and ensure the vendor you choose has a secure and suitable option for your business needs. Depending on your environment and corporate objectives, you may want to opt for 100% cloud deployment, an on-premise installation, or a hybrid option connecting in-house enterprise and cloud deployments to allow more flexibility. Cloud connections should use the HTTPS protocol with user authentication, and HIPAA-compliant servers should be available where required.

**6****Flexibility**

An efficient NLP solution will be easily adaptable to different solution areas, so that users can get immediate answers across a broad range of questions, with rapid refinement and iteration to get to the required answer. The system should handle content sourced both internally and externally, and be able to leverage in-house ontologies.

Ideally, you want out-of-the-box capabilities to ensure you can get up and running quickly, while also being able to create your own searches.

**7****Usability**

You want an NLP solution that is accessible to both power-users and less experienced users, including options to provide broad access to non-technical users. Look for a solution that offers multiple interfaces, plus the ability to customize interfaces to your exact needs via something like a web portal.

**8****Integration with machine learning engines**

ML is increasingly being used to help solve complex issues by analyzing data from textual sources. A suitable NLP solution should be able to process the mass of unstructured data in electronic health records, clinical trial records or full-text literature, to provide the clean, well-structured data that is needed to drive a predictive modeling ML engine. A standard API should also allow the NLP solution to be plugged into the selected ML tool.

**9****Transparency**

A successful NLP solution should be trusted by its users, which requires transparency rather than a black box. The results should be understood, and they should be reproducible and predictable. The system should be standards-based, open and auditable, and offer query quality assessment via Gold Standard evaluation.

**10****Regular updating, references and use cases**

Any NLP solution you might consider needs to be regularly updated, so look at the product release cycle of the platform and how information is shared. Check who else in your industry is using the solution, look at the vendor's case studies, and check for published examples in peer-reviewed journals. Any good NLP solution should have clear testimonials and relevant use cases for those in the healthcare, pharmaceutical and biotechnology spaces.

The success of any natural language processing solution will be in its ability to deliver quantifiable Return on Investment. Ultimately, the goal is for you to spend less time doing manual work and ensure that you make the most of your text, to get you the answers you need.

**For more information on the Linguamatics NLP platform, please contact us at: [nlp@iqvia.com](mailto:nlp@iqvia.com)**