Strix - Språkbanken's Text Analysis Platform

User Manual

Språkbanken Text

May 22, 2025

Contents

Introduction

Strix is Språkbanken's cutting-edge platform for advanced text analysis and exploration. Strix empowers researchers, linguists, and organizations to analyze diverse datasets, uncover patterns, and gain insights into textual data. Whether you're working with historical texts, political speeches, or modern corpora, Strix provides the tools you need to explore, visualize, and understand your data.

This guide will walk you through the features and functionalities of Strix, from performing simple searches to visualizing metadata and exploring semantic relationships between documents. With Strix, you can harness the power of language technology to unlock the full potential of your datasets.

Overview of Strix documentation

Here's what you can find in this documentation:

- **Quick start guide**: A step-by-step guide to help you get started with Strix quickly.
- **What is Strix?**: An introduction to Strix, its purpose, and its key features.
- **Data selection**: Learn more about the data in Strix, including modes, corpora, and corpus details.
- **Search**: Understand the two search formats in Strix: simple search and document search.
- **Filters**: Learn how to narrow down your search results using filters.
- **Data visualization**: Explore how to view a list of documents with their previews, analyze metadata using statistics, and visualize geo-locations on interactive maps.
- **Document view**: Dive into document reader editor to read the full document and analyze metadata on word level using statistics.
- **Related documents**: Discover similar documents to the one in focus.
- **Login access**: How to gain access to Strix.

1 Quick Start Guide

Welcome to the **Strix quick start guide**! This guide provides step-by-step instructions to help you get started with Strix quickly. Below are the key tasks you can perform:

1. **Search for documents** Quickly find relevant documents using simple or advanced search options.

2. **Document view** Explore individual documents in detail, including their content and metadata.

3. **Select corpora** Choose specific datasets to focus your analysis.

4. **Switch modes** Explore different modes like Modern, Mink, or Parallel and how to switch between modes.

5. **Explore related documents** Discover semantically similar documents to uncover deeper insights.

6. **Statistics and maps** Visualize metadata and geo-locations to analyze patterns and trends.

7. **Adding your own data to Strix** Upload and analyze your custom datasets with advanced tools.

1.1 Search for Documents

Imagine you're curious about how different political parties in Sweden approach the topic of **klimat politik** (climate policy). Strix can help you uncover insights by searching through political manifestos, speeches, and other documents. Let's explore how you can use Strix to dive into this topic.

1.1.1 Simple search: Starting with a word

You decide to start with a simple question: *What do political documents say about "klimat"?*

1. Navigate to the **Search bar** at the top of the Strix interface.

2. Type the word `klimat` into the search bar.

3. Press the **Search** button or hit `Enter`.

4. Strix will return a list of documents mentioning the word "klimat," allowing you to explore how it is discussed across different contexts.

Example: Search for the word `klimat` in the **Swedish party programs and election manifestos** corpus. Search example: klimat

1.1.2 Document search: Exploring climate policy

Now, you want to dig deeper. You're interested in understanding how **klimat politik** is framed by different political parties. Strix's **Document search** feature can help you find semantically similar documents that discuss this topic.

1. Select the **Document search** tab (if available).
2. Navigate to the **Search bar**.
3. Enter the phrase `klimat politik` into the search bar.
4. Press the **Search** button or hit `Enter`.
5. Strix will retrieve the top 50 documents that are semantically similar to your query. These documents may include political manifestos, speeches, or reports that discuss climate policy.

Example: Search for the phrase `klimat politik`. Search example: klimat politik

1.1.3 Related documents – What do the parties think?

You've gathered some insights, but now you want to compare how different political parties approach **klimat politik**. Strix's **Related documents** feature can help you connect the dots and uncover deeper relationships between documents.

1. Select a document from the search results that seems particularly interesting (e.g., a manifesto from a specific party).
2. Click on the **Related documents** button to explore other documents that are semantically similar.

Example: Explore documents related to a political manifesto in the **Swedish party programs and election manifestos** corpus. This will help you see how different parties frame their stance on **klimat politik** and identify recurring themes or contrasting viewpoints.

*For instance, you might find that one party focuses on **sustainability**, while another emphasizes **economic growth**. The **Related documents** feature allows you to compare these perspectives side by side, helping you build a more comprehensive understanding of the discourse.*

1.1.4 Conclusions

By following these steps:

- You started with a broad search to understand the general discourse.
- You narrowed your focus to explore specific policies and stances.
- You used related documents and visualizations to compare perspectives across parties.

Strix empowers you to uncover insights and build a comprehensive understanding of how **klimat politik** is discussed in Swedish politics. Now it's your turn to explore further and uncover deeper insights hidden in the data.

1.2 Document View

After performing a search, you can explore individual documents in detail using the **Document View** feature. This section allows you to analyze the content, metadata, and linguistic attributes of a document.

1.2.1 Steps to use document view

1. Select a document from the search results.
2. The document will open in the **Document Reader**, where you can view its full content and metadata.
3. Use the **Annotations selector** to highlight specific linguistic features, such as verbs or nouns.
4. Switch to the **Statistics tab** to analyze word-level metadata attributes, such as part of speech or sentiment.

Example: Open a document from the **Swedish party programs and election manifestos** corpus to explore its content and metadata.

1.2.2 Highlighting annotations

Use the **Annotations selector** to highlight specific linguistic features, such as verbs or nouns. For an example of highlighting, see figure 11 in the Document View section.

1.2.3 Key features of document view

- **Full document display**: View the entire content of the document, including text and metadata.
- **Annotations and search**: Highlight specific linguistic features and search within the document.
- **Statistics tab**: Analyze word-level metadata attributes in a tabular format.
- **Mobile-friendly design**: Access the document view seamlessly on mobile devices.

This feature provides a comprehensive way to explore and analyze individual documents, helping you uncover deeper insights into your data.

1.3 Select Corpora

Corpora are collections (datasets) in Strix, each containing a number of documents. Follow these steps to select or deselect corpora:

1.3.1 Steps to select corpora

1. Navigate to the **Data selector** on the top-right side of the interface.
2. Use the checkboxes to select or deselect corpora.
3. The selected corpora will update the documents and filters dynamically.

Example: Select the **Swedish party programs and election manifestos** corpus to focus on political documents.

1.3.2 Buttons in the Data selector

- **Select all**: Selects all corpora in the current mode.
- **Deselect all**: Deselects all selected corpora.
- **Default**: Resets the selection to the default corpora for the current mode.

1.4 Switch Modes

Modes in Strix categorize datasets based on their type. Follow these steps to switch modes:

1.4.1 Steps to switch modes

1. Navigate to the **Mode selector** above the Strix logo on the top-left side of the interface.
2. Click on a mode (e.g., Modern, Mink, Parallel) to select it.
3. The selected mode will update the available corpora in the **Data selector**.

Example: Switch to the **Parallel** mode to explore datasets with source and reference documents, such as translations or OCR-corrected texts.

1.4.2 Default mode

The default mode in Strix is **Modern**, which contains datasets written in contemporary Swedish.

1.5 Explore Related documents

The **Related documents** feature helps you find documents that are semantically similar to a selected document. Follow these steps to explore related documents:

1.5.1 Steps to explore Related documents

1. Select a document from the search results or document list.
2. Click on the **Related documents** button.
3. View the list of semantically similar documents in the **Related documents** tab.

Example: Explore documents related to a political manifesto in the **Swedish party programs and election manifestos** corpus.

1.5.2 Graph visualization

1. Switch to the **Graph view** to visualize relationships between the selected document and its related documents.
2. Interact with the graph by zooming in/out or clicking on nodes to view more details.
3. **Graph view**: Available only for lexical datasets.

1.6 Statistics and Maps in Strix

Strix provides powerful tools for visualizing and analyzing data through **Statistics** and **Maps**. These features help you uncover patterns, trends, and relationships in your datasets.

1.6.1 Statistics: Analyze metadata attributes

The **Statistics** section allows you to explore metadata attributes at the text level and their elements, showing how many documents in a collection belong to each category.

How to use statistics:

1. Navigate to the **Statistics** tab in the Strix interface.
2. On the **left side**, select a metadata attribute (e.g., "Text classification (Blingbring)" or "Year").
3. On the **right side**, view the frequency of each element in the selected attribute across your chosen collections.

Example: Analyze year distribution

- Select the **Year** attribute from the metadata list.
- View the table on the right to see how many documents were published in each year.
- Click on a specific year (e.g., `1920`) to view all documents from that year.

Interactive features:

- **Click on frequency**: View documents from a particular dataset that match the element.
- **Click on element**: Click on a metadata element (e.g., a specific year or topic) to view documents where each document contains this element in its metadata.

1.6.2 Maps: Visualize geo-locations

The **Maps** section enables you to explore geographical data associated with your documents. Geo-locations mentioned in the documents are plotted on an interactive map.

How to use maps:

1. Navigate to the **Maps** tab in the Strix interface.
2. Select one or more collections to display geo-locations from those datasets.
3. Interact with the map:
   - **Click on points**: View detailed information about the documents mentioning a specific geo-location.
   - **Zoom in/out**: Explore clusters of geo-locations or focus on individual points.

Example: Explore geo-locations in political documents

- Select the **Swedish party programs and election manifestos** corpus.
- View the map to see locations mentioned in political manifestos.
- Click on a location (e.g., ”Stockholm”) to see all documents referencing it.

Interactive features:

- **Clusters**: Large datasets like Wikipedia are grouped into clusters for easier navigation. Zoom in to break clusters into individual points.
- **Documents**: Click on a geo-location to open a popover box that displays the number of documents mentioning this location. Clicking on the ”Show hits” button will display these documents in a tab located beside the Maps tab.

1.6.3 Why use statistics and maps?

- **Statistics**: Helps you analyze the distribution of metadata attributes, such as publication years, topics, or sentiment labels.
- **Maps**: Provides a spatial perspective, enabling you to explore geographical patterns and relationships in your data.

These tools make it easier to uncover insights and gain a deeper understanding of your datasets.

1.7 Adding your own text data to Strix

Strix allows users to upload their own text data (a collection of documents) and leverage its advanced functionalities to analyze, visualize, and explore the data. This feature is particularly useful for researchers, linguists, and organizations looking to exploit their custom datasets.

1.7.1 Why add your own data?

By adding your own text data to Strix through **Mink**, you can:

- **Perform advanced searches**: Use simple and document searches to explore your custom datasets.
- **Visualize metadata**: Analyze metadata attributes and geo-locations using **Statistics** and **Maps**.
- **Explore semantic relationships**: Use the **Related documents** feature to uncover connections between documents.
- **Analyze linguistic patterns**: Dive into word, sentence, and text-level metadata attributes for deeper insights.

1.7.2 What Is Mink?

Mink is Språkbanken's data platform that allows users to upload their collections and apply advanced language technology methods to their texts. The resulting annotated data can be:

- **Downloaded** for offline use.
- **Integrated** into research tools like **Korp** and **Strix** for further analysis.

You can read more about Mink and its documentation and tutorials at `https://spraakbanken.gu.se/en/tools/mink`. All data uploaded to Mink is securely stored behind a login and is not publicly available to other users.

1.7.3 How to add your data to Strix

Below are the steps to upload your data in Mink, annotate it, and make it available in Strix.

1. Prepare your data Before uploading your data, ensure it meets the following requirements:

- **File format**: Supported formats include `.txt`, `.docx`, `.odt`, `.pdf`, or `.xml`.
- **Metadata**: Include metadata for each document (e.g., title, author, year, genre) as tags/attributes if the file format is `.xml`.
- **Encoding**: Use UTF-8 encoding to ensure compatibility.
- **File size**: Ensure individual files do not exceed the maximum upload size (e.g., 10 MB per file).

2. Upload your data

1. Log in to the **Mink** platform.
2. Create a **corpus name** for your collection.
3. Select your files or drag and drop them into the upload area.
4. Edit the configuration if needed. By default, Mink creates a configuration for each corpus, which includes the following annotations added to each document using the **Sparv annotation tool**:
    - **Part of speech tags**
    - **Base form (Lemma)**
    - **Morphosyntactic tags (MSD)**
    - **Dependencies**
    - **Sentiment labels**
5. Run the annotation process.
6. Once the annotation is completed, the annotated data will be ready for download and available for installation in **Strix** and **Korp**.

3. Index your data and install in Strix After annotating your data, install the corpus into Strix:

1. Install the annotated corpus into Strix from the Mink platform.
2. Strix will automatically **index your data** to make it searchable and compatible with its advanced features.
3. Monitor the indexing progress in the **Status** section in Mink.
4. Once the installation is complete, the **Status** section will display a "Done" message.

4. Access your data in Strix After indexing, your data will appear under the **Mink mode** (personal collections) in Strix. You can either:

- Go to Strix and log in to view your data in **Mink mode**.
- Or, follow the link from Mink to Strix by clicking on the **Open** button located next to the **Install** button.

Once in Strix, you can:

- **Select your dataset** to perform searches and visualizations.
- **Combine your dataset** with other existing corpora for comparative analysis.

1.7.4 Example use case: Analyzing global warming

Imagine you are a researcher studying **global warming** and its representation in political speeches. You have a collection of speeches and reports that you want to analyze. Here's how you can use Mink and Strix to explore your data:

1. **Upload your data**:
   - Prepare your collection of speeches and upload them to Mink.
   - Annotate the data using Sparv to add linguistic metadata like part of speech tags and sentiment labels.
2. **Install in Strix**:
   - Install the annotated corpus into Strix and index it.
3. **Perform searches**:
   - Use **Simple search** to find occurrences of terms like `global warming (global uppvärmning)` or `climate change (klimatförändring)`.
   - Use **Document search** to explore semantically similar documents discussing renewable energy or sustainability.
4. **Visualize metadata**:
   - Use the **Statistics** tab to analyze the frequency of terms like "carbon emissions" or "renewable energy."
   - Use the **Maps** tab to visualize geo-locations mentioned in the speeches, such as references to international climate agreements.
5. **Explore related documents**:
   - Use the **Related documents** feature to find connections between speeches from different political parties or organizations.

By following these steps, you can uncover patterns, trends, and insights into how global warming is discussed in your dataset.

1.7.5 Troubleshooting and support

If you encounter any issues while uploading or indexing your data:

- Ensure that your files meet the format and size requirements.
- Check the **Status** section in Mink for error messages or warnings.
- Contact the Strix support team at sb-info@svenska.gu.se for assistance.

Start uploading your data today and unlock the full potential of Strix for your research!

2 What is Strix?

Strix is Språkbanken's text analysis platform, designed for advanced research and exploration of textual data. It is similar to Korp, Språkbanken's word research platform for searching large amounts of text. However, Strix focuses on full text and a broader range of text analysis capabilities.

The data in Strix is highly diverse, including sources such as newspapers, novels, governmental data, Wikipedia, historical texts, and much more. Each dataset is referred to as a **corpus**, and each corpus contains a collection of text documents. These documents have been annotated by Språkbanken's analysis platform, Sparv, at the word, sentence, and text levels.

2.1 Key Features of Strix

- **Document View**: View the content of each document and its annotations generated by Sparv (from word to text level) in Codemirror editor view.
- **Document Statistics**: Analyze token-level statistics for various word attributes within a document.
- **Search Capabilities**:
  - Search within individual documents.
  - Perform simple searches or document searches across a selected collection of corpora.
- **Data Visualization**: View statistics and graphs for selected corpora and documents, and explore text attributes connected to selected documents.
- **Maps Section**: Visualize locations mentioned in documents on an interactive map.
- **Related Documents**: Explore similar documents using document vector search powered by KBLab's KB-SBERT sentence transformers.
- **Filters**: Narrow down your search results using advanced filters.
- **Sparv Integration**: Visualize and search in those analyzes produced by Sparv.
- **And More**: Strix offers many additional features to enhance your text analysis experience.

To make navigation easier, the Strix documentation includes images alongside text, helping users understand the platform's features and functionality.

3 Data Selection

Strix contains a diverse collection of corpora (datasets or documents) ranging from historical to modern data. Some datasets in Strix are open access and can be viewed without restrictions, while others require login access. Each corpus provides a unique perspective, allowing users to explore and analyze textual data in detail.

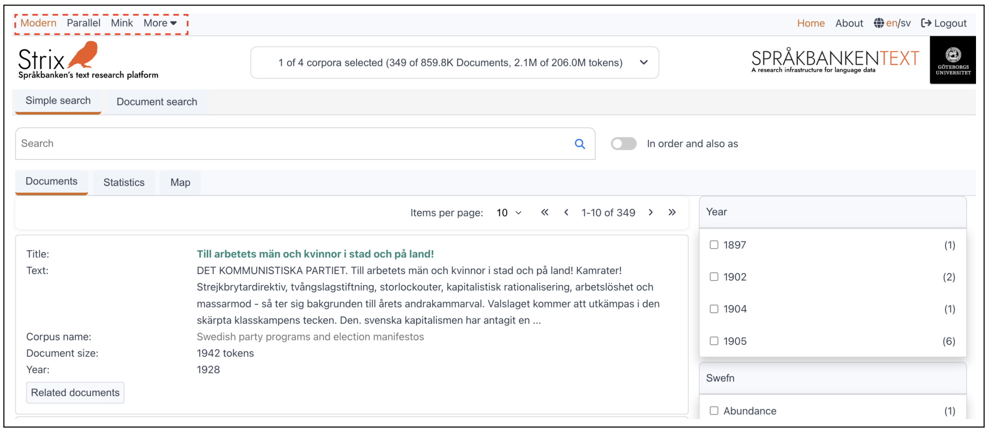
Each corpus in Strix belongs to one or more **modes**. Modes are created based on the type of collection. For example:

- Data from 1900 to the present is categorized under the **Modern** mode.
- The **Mink** mode is available for users who are logged in and have personal collections in Strix. This mode allows users to access and analyze their private datasets securely.
- The **Parallel** mode is designed for datasets where each document has a corresponding reference document. This mode is useful for tasks like translation alignment, OCR correction, or comparing student essays with teacher corrections.
- Other modes are created based on specific collections.

More details about modes can be found in the Modes subsection below.

3.1 Modes

The datasets in Strix are divided into different **modes**, such as Modern, Parallel, Mink, and many others. These modes are accessible on the Strix platform, located right above the Strix logo on the top-left side, as shown in the figure below. The default mode in Strix is the **Modern** mode.



1 of 4 corpora selected (349 of 859.8K Documents, 2.1M of 206.0M tokens)

In order and also as

Items per page: 10 « ‹ 1-10 of 349 › »

Title: Till arbetets män och kvinnor i stad och på land!

Text: DET KOMMUNISTISKA PARTIET. Till arbetets män och kvinnor i stad och på land! Kamrater! Strejkbrytardirektiv, tvångslagstiftning, storlockouter, kapitalistisk rationalisering, arbetslöshet och massarmod - så ter sig bakgrunden till årets andrakammarval. Valslaget kommer att utkämpas i den skärpta klasskampens tecken. Den. svenska kapitalismen har antagit en ...

Corpus name: Swedish party programs and election manifestos

Document size: 1942 tokens

Year: 1928

Year

□ 1897	(1)
□ 1902	(2)
□ 1904	(1)
□ 1905	(6)

Swefn

□ Abundance	(1)

Figure 1: Modes in Strix.

The selected mode is always highlighted with a distinct color. Once a mode is selected, the corpora in the **Corpora Selection** section are updated to reflect the corpora available in the selected mode. More details about corpora selection can be found in the Corpora section.

Below is a list of modes available in Strix, along with their descriptions and examples:

Modes in Strix

- **Modern**: This mode contains datasets written in contemporary Swedish (from the 1900s to the present). The datasets in this mode are open access. **Examples of datasets in Modern mode**:
  - Swedish party programs and election manifestos
  - Swedish Wikipedia
  - Riksdag open data (governmental)
- **Mink**: This mode is only available if the user is logged into Strix and has one or more personal collections in Strix. It is a protected mode and is not visible to users who are not logged in.
- **Parallel**: The Parallel mode is unique compared to other modes. In this mode, each document has a corresponding reference document. When a user opens a document in this mode, the Codemirror editor displays two documents side by side:
  - The **source document**.
  - The **reference document** (linked to the source document).

  Why two documents? The datasets in Parallel mode often involve translations or corrections. Examples include:
  - Translations from one language to another (e.g., novels, Bible texts).
  - OCR-scanned documents normalized using NLP models to correct OCR errors.
  - Handwritten essays by students learning Swedish, where:
    - The **source text** is written by the student.
    - The **target text** is normalized and corrected by the teacher.

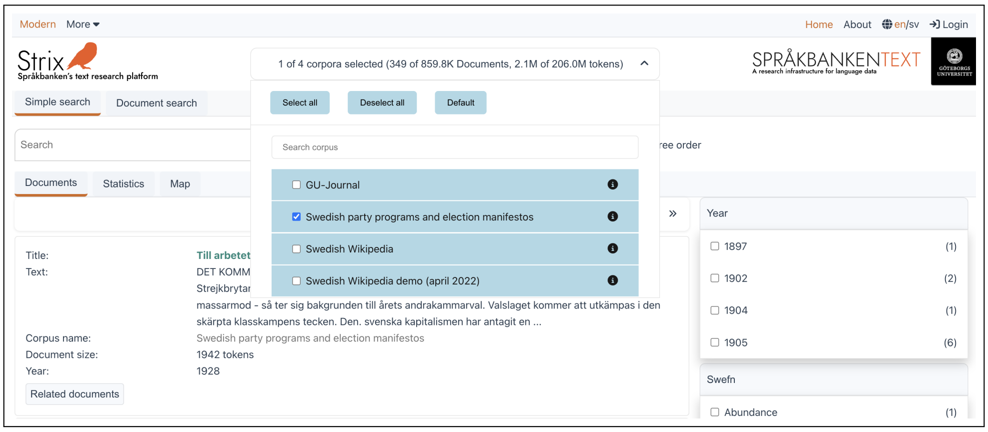
  Examples of datasets in Parallel mode:
  - Translated novels
  - Bible texts
  - OCR-corrected documents
  - Student essays with teacher corrections
- **More**: The **More** button, located on the far right, is a dropdown menu containing additional modes in Strix. These include:
  - Detektiva avdelningen
  - Jubilee Archive
  - The Swedish Literature Bank

For detailed information about corpora in each mode and how to select them, see the Corpora section.

3.2 Corpora

The **Data selector** (or **Corpus selector**) is used to choose one or more corpora. Users can find this feature right beside the Strix logo on the top-right side of the platform. When a mode is selected, the default corpora for that mode are automatically selected. However, users can customize their selection by selecting or deselecting corpora based on their preferences and needs.

Every time a corpus is selected or deselected in the **Data selector** (as shown in the image below), the documents and tables in the **Filter section** (on the right side) are updated accordingly. More details about the filter section will be covered later.



1 of 4 corpora selected (349 of 859.8K Documents, 2.1M of 206.0M tokens)

- ☐ GU-Journal
- ☑ Swedish party programs and election manifestos
- ☐ Swedish Wikipedia
- ☐ Swedish Wikipedia demo (april 2022)

Title:	Till arbetet
Text:	DET KOMM Strejkbrytar massarmod - så ter sig bakgrunden till årets andrakammarval. Valslaget kommer att utkämpas i den skärpta klasskampens tecken. Den. svenska kapitalismen har antagit en ...
Corpus name:	Swedish party programs and election manifestos
Document size:	1942 tokens
Year:	1928

Year	
☐ 1897	(1)
☐ 1902	(2)
☐ 1904	(1)
☐ 1905	(6)

Swefn	
☐ Abundance	(1)

Figure 2: Data selector in Strix.

Buttons in the Data selector

The **Data selector** includes the following buttons to make selection easier:

- **Select all**: Selects all the corpora in the current mode.
- **Deselect all**: Deselects all the selected corpora.
- **Default**: Resets the selection to the default corpora for the currently selected mode.

Searching for Corpora

If the list of corpora is long, users can use the **Search** feature in the **Data selector** to quickly find the corpus they are looking for. This makes it easier to sort through large collections of corpora.

Corpus Information

Each corpus in the **Data selector** has an **info icon** button located on the right side of the corpus name. Clicking this button opens a dialog box that provides a detailed description of the corpus. More information about corpus details can be found in the Corpus Description section.

3.3 Corpus Description

Each corpus in Strix has a default metadata structure. Some corpora may also include additional annotations at the word and text levels. Below is an example of the basic structure of a corpus, using the **Swedish party programs and election manifestos** corpus as a reference.

Swedish party programs and election manifestos

- **Mode:** Modern
- **Documents:** 349
- **Corpus Size:** 2,099,602 tokens
- **Word attributes:**
  - Lemgram
  - Sense
  - Compound word forms
  - Compound lemgrams
  - Dependency relation
  - Dephead
  - Ref
  - Sentiment label
  - Text classification (blingbring)
  - Text classification (swefn)
  - Baseform
  - Msd
  - Part-of-speech
- **Text attributes:**
  - Text classification (blingbring)
  - Text classification (swefn)
  - Readability measure (LIX)
  - Readability measure (ovix)
  - Readability measure (nk)
  - Id
  - Party
  - Type
  - Year
- **Structural attributes:**
  - Name tag:
    - Expression
    - Name
    - Type
    - Subtype
  - Sentence
  - Location

This metadata structure provides a comprehensive overview of each corpus in Strix, enabling users to perform detailed analyses at both the word and text levels. For more information about how to use these attributes, refer to the relevant sections in the documentation.

4 Search

The search functionality in Strix is divided into two parts: **Simple search** and **Document search**.

- **Simple search**: Simple Search allows users to search for specific words or word forms. It also supports searching for expressions or phrases. More details about Simple Search can be found on the Simple Search page.
- **Document search**: Document Search uses vector search techniques to find vectors that are semantically close to the given query vector. The query vector can be a word, sentence, or document. More details about Document Search are explained in the Document Search section.

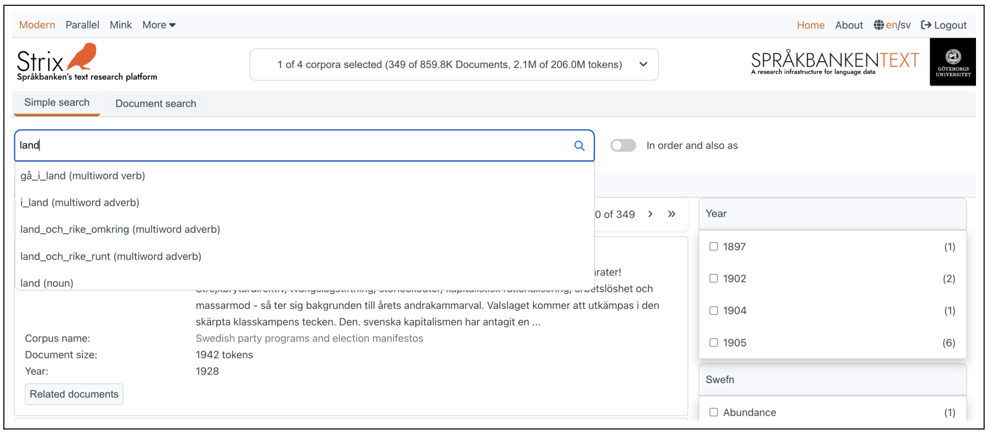
4.1 Simple Search

Simple Search allows users to search for an exact word, word form, or phrase. The resulting documents from the search are displayed in the section below the search bar. Let's explore the different functionalities of Simple Search.

4.1.1 Search for a word

This is a basic search where users can type a word and press the search button to retrieve documents. The search highlights the exact word entered in the documents.

Example: Searching for the word `klimat` in the **Swedish party programs and election manifestos** corpus. Search example: klimat



1 of 4 corpora selected (349 of 859.8K Documents, 2.1M of 206.0M tokens)

land

gå_i_land (multiword verb)
i_land (multiword adverb)
land_och_rike_omkring (multiword adverb)
land_och_rike_runt (multiword adverb)
land (noun)

rater!
etslöshet och massarmod - så ter sig bakgrunden till årets andrakammarval. Valslaget kommer att utkämpas i den skärpta klasskampens tecken. Den. svenska kapitalismen har antagit en ...

Corpus name: Swedish party programs and election manifestos
Document size: 1942 tokens
Year: 1928

Year	
□ 1897	(1)
□ 1902	(2)
□ 1904	(1)
□ 1905	(6)

Swefn	
□ Abundance	(1)

Figure 3: Search for a word form.

4.1.2 Search for a word form

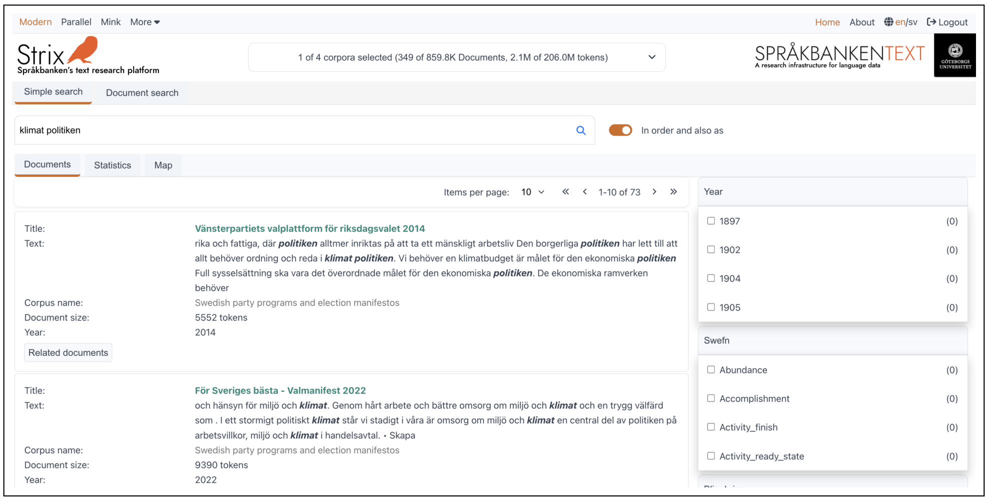
Instead of searching for an exact word, users can search for a word form (e.g., lemma or lemgram). As users start typing, the query is sent to the **Karp API**, which returns lemgrams for the input word. These lemgrams are displayed in a dropdown below the input field (see figure 3), allowing users to select a word form and search for it.

Example: Searching for the word form `land (noun)` in the **Swedish party programs and election manifestos** corpus. Search example: land (noun)

4.1.3 Search for an exact phrase or words in a phrase

Users can also search for a phrase instead of a single word or word form. To enable phrase search, users need to activate the **toggle button** located to the right of the search input field. This allows searching for an exact phrase or specific words within the phrase.

Example: Searching for the phrase `"klimat politiken"` in the **Swedish party programs and election manifestos** corpus. Search example: ”klimat politiken”



Title: Vänsterpartiets valplattform för riksdagsvalet 2014
Text: rika och fattiga, där ***politiken*** alltmer inriktas på att ta ett mänskligt arbetsliv Den borgerliga ***politiken*** har lett till att allt behöver ordning och reda i ***klimat politiken***. Vi behöver en klimatbudget är målet för den ekonomiska ***politiken*** Full sysselsättning ska vara det överordnade målet för den ekonomiska ***politiken***. De ekonomiska ramverken behöver
Corpus name: Swedish party programs and election manifestos
Document size: 5552 tokens
Year: 2014

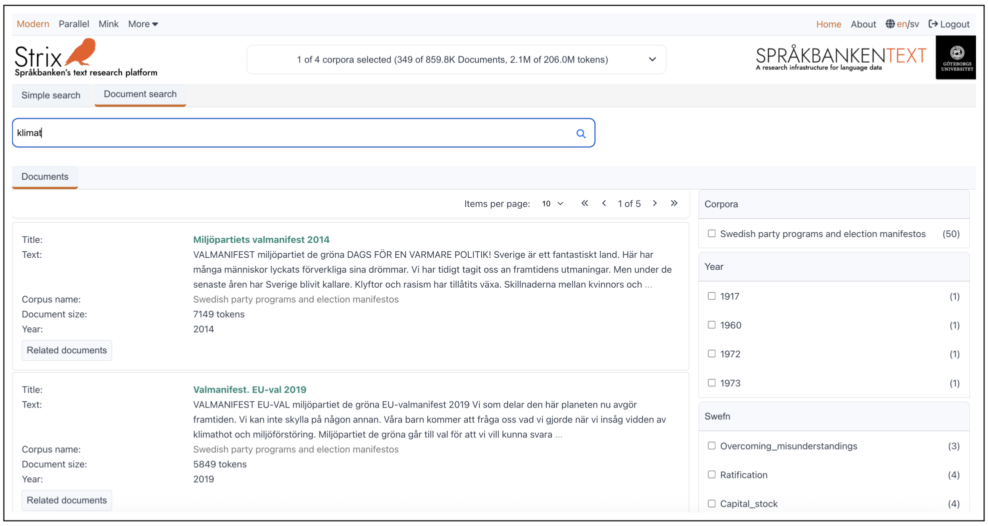
Title: För Sveriges bästa - Valmanifest 2022
Text: och hänsyn för miljö och ***klimat***. Genom hårt arbete och bättre omsorg om miljö och ***klimat*** och en trygg välfärd som . I ett stormigt politiskt ***klimat*** står vi stadigt i våra är omsorg om miljö och ***klimat*** en central del av politiken på arbetsvillkor, miljö och ***klimat*** i handelsavtal. • Skapa
Corpus name: Swedish party programs and election manifestos
Document size: 9390 tokens
Year: 2022

Figure 4: Exact phrase or words in a phrase.

4.2 Document Search

Every document in Strix has a document vector. These vectors are used in the document search functionality. At search time, the search query is converted into a vector and compared to the document vectors. The fifty closest documents to the query are returned.

These documents are the ones that are semantically close to the given vector query, as shown in the figure below. The current default number of documents that the document search returns is limited to 50, but this number will be a dynamic input instead.



Title: Miljöpartiets valmanifest 2014
Text: VALMANIFEST miljöpartiet de gröna DAGS FÖR EN VARMARE POLITIK! Sverige är ett fantastiskt land. Här har många människor lyckats förverkliga sina drömmar. Vi har tidigt tagit oss an framtidens utmaningar. Men under de senaste åren har Sverige blivit kallare. Klyftor och rasism har tillåtits växa. Skillnaderna mellan kvinnors och ...
Corpus name: Swedish party programs and election manifestos
Document size: 7149 tokens
Year: 2014

Title: Valmanifest. EU-val 2019
Text: VALMANIFEST EU-VAL miljöpartiet de gröna EU-valmanifest 2019 Vi som delar den här planeten nu avgör framtiden. Vi kan inte skylla på någon annan. Våra barn kommer att fråga oss vad vi gjorde när vi insåg vidden av klimathot och miljöförstöring. Miljöpartiet de gröna går till val för att vi vill kunna svara ...
Corpus name: Swedish party programs and election manifestos
Document size: 5849 tokens
Year: 2019

Figure 5: Document search result for word *klimat*.

KBLab's KB-SBERT is used to create the vectors and also to perform the document search. This means that the search does not look for exact matches of the query but instead finds documents that are semantically similar to the query based on vector representations.

Users can search for a word, phrase, sentence, or even a whole document. Below are some examples:

Examples

1. **Word search**

   Query: klimat

   Result: Documents in Swedish party programs and election manifestos that are semantically related to the word `klimat`.

2. **Phrase search**

   Query: klimat politik

   Result: Documents in Swedish party programs and election manifestos that are semantically related to the phrase `klimat politik`.

3. **Sentence search**

   Query: Våra barn kommer att fråga oss vad vi gjorde när vi insåg vidden av klimathot och miljöförstöring

   Result: Documents in Swedish party programs and election manifestos that are semantically similar to the sentence.

4. **Document search**

   Query: A full document text.

   Result: Documents with content or context that is semantically similar to the provided document.

5 Filters

Filters in Strix are one of the core functionalities, playing a crucial role in narrowing down search results. When working with a large collection of documents from various genres, such as newspapers, historical texts, and more, filters allow users to refine their search queries and focus on specific subsets of documents based on predefined criteria.

5.1 How Filters Work

Filters in Strix are designed to support advanced and complex filtering capabilities. Users can scroll through the available options in each metadata filter to refine their search. Here's how it works:

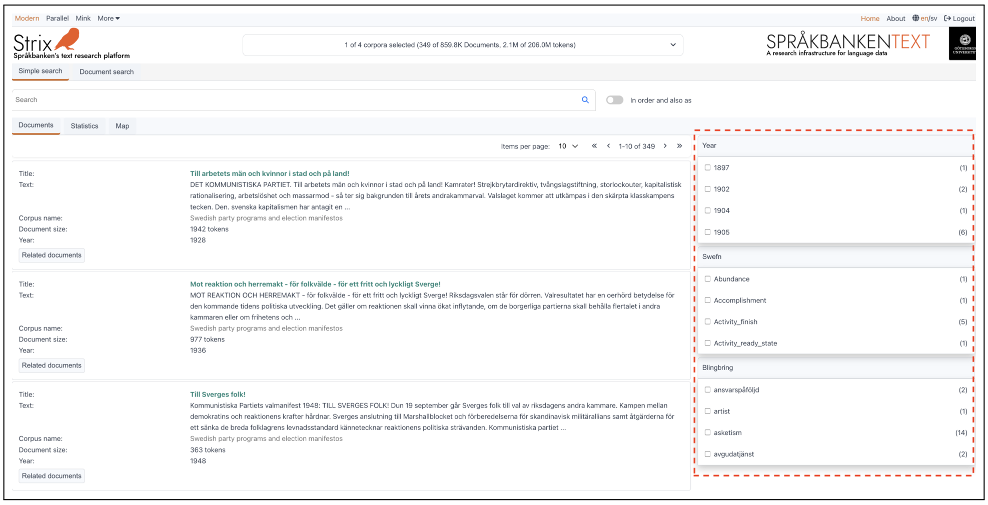
1. **Indexing metadata** Each document in the collection is indexed with metadata at three levels:
   - **Text level**: Metadata such as genre, newspaper, year, author, topics, and more.
   - **Sentence level**: Metadata such as named entities and geo-locations.
   - **Word level**: Metadata such as part of speech, word form, sentiment analysis, and more. (More details about word-level metadata can be found in the Document View section.)
2. **Applying filters** When a user applies a filter (e.g., selecting "year," "topics," or other metadata), Strix uses the indexed metadata to narrow down the search results to documents that match the filter criteria.
3. **Combining filters** Users can combine multiple filters to refine their search further. For example, they can filter for "newspapers" published in 1905.
4. **Efficient query execution** The filtering process is optimized to ensure that filters are applied quickly and efficiently. This allows users to refine their searches seamlessly, even when working with large datasets like Wikipedia, which contains more than 800,000 documents in the Swedish language.

5.2 Examples of Filters

1. **Year filter** Focus on documents from a specific year, such as "1920." Search Example: Year 1920 *(Example corpus: Swedish Party Programs and Election Manifestos)*
2. **Text classification (SweFN)** Retrieve documents with a specific topic, such as "Satisfying." Search example: SweFN topic - Satisfying *(Example corpus: Swedish Party Programs and Election Manifestos)*
3. **Text classification (Blingbring)** Search for documents with a specific Blingbring topic, such as "afton." Search example: Blingbring topic - Afton *(Example corpus: Detektivaavdelningen)*

5.3 Standard Filters

Currently, the **Year**, **Text classification (SweFN)**, and **Text classification (Blingbring)** filters are available on the right-hand side of the interface, as shown in the figure below. These are referred to as **Standard filters** and provide quick access to commonly used filtering options.



1 of 4 corpora selected (349 of 859.8K Documents, 2.1M of 206.0M tokens)

Items per page: 10 « ‹ 1-10 of 349 › »

Title: Till arbetets män och kvinnor i stad och på land!
Text: DET KOMMUNISTISKA PARTIET. Till arbetets män och kvinnor i stad och på land! Kamrater! Strejkbrytardirektiv, tvångslagstiftning, storlockouter, kapitalistisk rationalisering, arbetslöshet och massarmod - så ter sig bakgrunden till årets andrakammarval. Valslaget kommer att utkämpas i den skärpta klasskampens tecken. Den. svenska kapitalismen har antagit en ...
Corpus name: Swedish party programs and election manifestos
Document size: 1942 tokens
Year: 1928

Title: Mot reaktion och herremakt - för folkvälde - för ett fritt och lyckligt Sverge!
Text: MOT REAKTION OCH HERREMAKT - för folkvälde - för ett fritt och lyckligt Sverge! Riksdagsvalen står för dörren. Valresultatet har en oerhörd betydelse för den kommande tidens politiska utveckling. Det gäller om reaktionen skall vinna ökat inflytande, om de borgerliga partierna skall behålla flertalet i andra kammaren eller om frihetens och ...
Corpus name: Swedish party programs and election manifestos
Document size: 977 tokens
Year: 1936

Title: Till Sverges folk!
Text: Kommunistiska Partiets valmanifest 1948: TILL SVERGES FOLK! Dun 19 september går Sverges folk till val av riksdagens andra kammare. Kampen mellan demokratins och reaktionens krafter hårdnar. Sverges anslutning till Marshallblocket och förberedelserna för skandinavisk militärallians samt åtgärderna för ett sänka de breda folklagrens levnadsstandard kännetecknar reaktionens politiska strävanden. Kommunistiska partiet ...
Corpus name: Swedish party programs and election manifestos
Document size: 363 tokens
Year: 1948

Year
- □ 1897 (1)
- □ 1902 (2)
- □ 1904 (1)
- □ 1905 (6)

Swefn
- □ Abundance (1)
- □ Accomplishment (1)
- □ Activity_finish (5)
- □ Activity_ready_state (1)

Blingbring
- □ ansvarspåföljd (2)
- □ artist (1)
- □ asketism (14)
- □ avgudatjänst (2)

Figure 6: Standard filters in Strix.

5.4 Advanced Filters

For **Advanced search**, all indexed metadata will be available as filtering options. Since each collection contains a vast amount of metadata, it is challenging to fit all options on the main page. **Advanced filters** provide a more comprehensive filtering experience, allowing users to refine their search using the full range of metadata attributes.

Filters empower users to explore and analyze large collections of documents effectively, making it easier to derive insights and find relevant information.

6 Data Visualization

Data visualization in Strix provides users with powerful tools to explore and analyze large collections of documents in an intuitive and interactive way. It is divided into three main sections:

1. **Documents** The documents are shown with a preview only, allowing users to quickly scan the content. When a user clicks on a document, the full document opens, providing detailed insights into its structure, semantics, and key information.

2. **Statistics** The statistics section currently displays data in tabular format, helping users understand the distribution and frequency of metadata like genres, publication years, authors, and more. While graphs and charts are not available yet, they will be introduced in future updates to provide visual summaries and make data analysis even more intuitive.

3. **Maps** The maps section enables users to visualize geographical data associated with the documents. By plotting named entities or geo-locations on a map, users can explore spatial patterns and relationships within the dataset.

Each of these sections is designed to provide a unique perspective on the data, making it easier to uncover insights and gain a deeper understanding of the information in the Strix collections.

Explore the subsections to learn more about how each visualization tool works and how it can help you analyze your data effectively.

6.1 Documents

This section is a collection of documents, similar to how Google displays search results. When users search in Strix, they get a list of documents from the selected collections. The documents are shown with a preview only, allowing users to quickly scan the content. When a user clicks on a document, the full document opens, providing detailed insights into its structure, semantics, and key information.

Each document in the collection is displayed as shown in the figure below. Here's what users can expect to see for each document:

1. **Title** The title of the document. A quick glance at the title gives users an idea of the document's content.

2. **Text** A preview of the text in the document (usually the first 50 tokens). This snippet helps users decide if the document is relevant to their search.

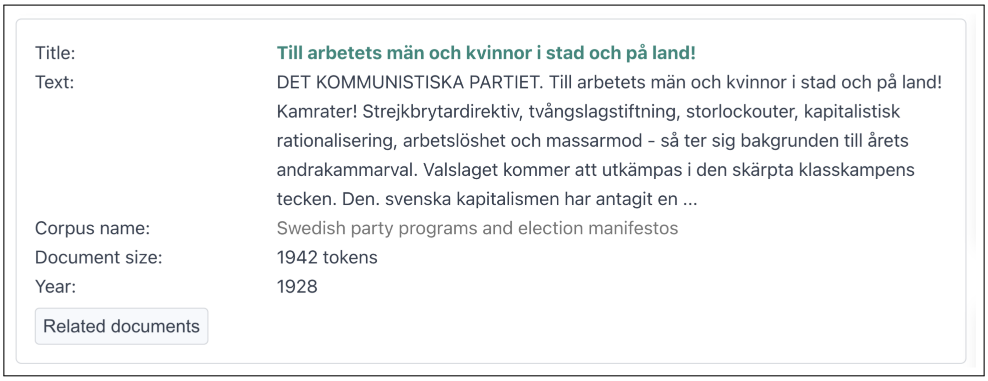
3. **Corpus name** The name of the collection that the document belongs to. This helps users identify the source of the document.

4. **Document size** The number of tokens (words or word-like units) in the document. A handy detail for understanding the document's length.

5. **Year** The year the document was created, based on metadata. Note: Some documents may not have year information if it's missing in the metadata.

6. **Related documents** A button that opens a tab right beside the **Maps** section. This tab displays the top 50 other documents in the collection that are semantically close to the current document. Perfect for exploring similar content!

7. **Link** Some collections provide a link to the source of the document. If a URL is available in the metadata, it will be displayed here for easy access.



Title: Till arbetets män och kvinnor i stad och på land!

Text: DET KOMMUNISTISKA PARTIET. Till arbetets män och kvinnor i stad och på land! Kamrater! Strejkbrytardirektiv, tvångslagstiftning, storlockouter, kapitalistisk rationalisering, arbetslöshet och massarmod - så ter sig bakgrunden till årets andrakammarval. Valslaget kommer att utkämpas i den skärpta klasskampens tecken. Den. svenska kapitalismen har antagit en ...

Corpus name: Swedish party programs and election manifestos

Document size: 1942 tokens

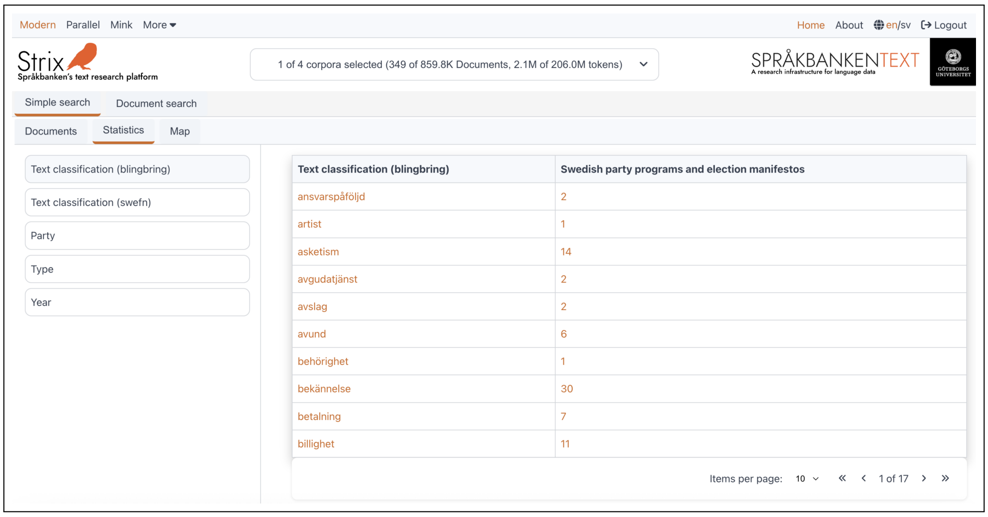
Year: 1928

Figure 7: Each document hit in Strix.

This layout ensures that users can quickly scan and interact with the documents, making it easier to find relevant information and explore related content. Dive in and discover the power of Strix's document visualization!

6.2 Statistics

The **Statistics** section in Strix provides users with the ability to explore metadata attributes and their elements, showing how many documents in a collection belong to a particular element. The statistics view is divided into two parts, as shown in the figure below:



1 of 4 corpora selected (349 of 859.8K Documents, 2.1M of 206.0M tokens)

Text classification (blingbring)	Swedish party programs and election manifestos
ansvarspåföljd	2
artist	1
asketism	14
avgudatjänst	2
avslag	2
avund	6
behörighet	1
bekännelse	30
betalning	7
billighet	11

Figure 8: Statistics for metadata attribute *Blingbring*.

On the **left side**, users can see the metadata attributes available in the selected collections. On the **right side**, a tabular view displays the statistics for the selected metadata attribute. Below is a detailed explanation of these two parts and how they work.

6.2.1 Metadata section

The **left side** of the statistics view contains a list of metadata attributes. This list updates dynamically whenever a collection is selected or deselected. The list represents the **union** of metadata attributes available across the selected collections.

- By default, the metadata attribute **"Text classification (Blingbring)"** is selected when the user navigates to the statistics page.
- The table on the right updates automatically whenever a new metadata attribute is selected from this list.

This dynamic behavior ensures that users always see the most relevant metadata attributes for their selected collections.

6.2.2 Table view

The **right side** of the statistics view displays a table with the statistics for the currently selected metadata attribute. The table is structured as follows:

1. **First column** This column lists the **elements** of the currently selected metadata attribute (e.g., elements in "Blingbring" as shown in the figure 8). These elements update dynamically whenever the user selects or deselects a collection.
2. **Second column** This column represents the **first collection** that the user selected. It shows the frequency of each element in that collection.
3. **Dynamic columns** Columns beyond the second are added or removed dynamically based on the user's selection or deselection of collections. Each column corresponds to a selected collection and shows the frequency of the elements in that collection.

6.2.3 Interactive features

Each value and frequency in the table is color-coded for interactivity:

- **Black text**: Indicates that the value is **not clickable**. This occurs when an element in the selected metadata attribute has a frequency of `0`.
- **Colored text**: Indicates that the value is **clickable**. Clicking on these values opens a new tab right after the **Maps** section, displaying the documents associated with the selected element or frequency.

Click behavior:

- **Clicking on an element**: Displays all documents across the selected collections that contain the element.
- **Clicking on a frequency**: Displays the documents from the specific collection that contain the element with the selected frequency.

This intuitive design allows users to explore metadata attributes and their elements in detail, making it easier to analyze and navigate large collections of documents.

6.3 Maps

The **Maps** section in Strix enables users to visualize geographical data associated with the documents. By plotting geo-locations on a map, users can explore spatial patterns and relationships within the dataset. The map dynamically updates every time the user selects one or more collections, displaying the geo-locations mentioned in the documents from the selected collections (see figure 9 for an example visualization).

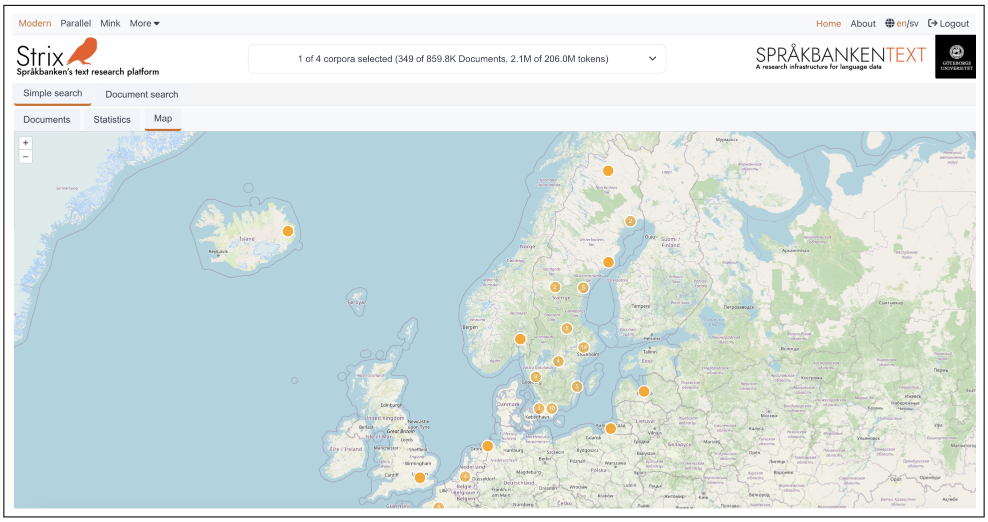

Modern
Parallel
Mink
More
Home
About
en/sv
Logout
Strix
Språkbanken's text research platform
1 of 4 corpora selected (349 of 859.8K Documents, 2.1M of 206.0M tokens)
SPRÅKBANKENTEXT
A research infrastructure for language data
Simple search
Document search
Documents
Statistics
Map
+
−

Figure 9: Geo-locations mentioned in *Swedish party programs and election manifestos* corpus.

6.3.1 Key features

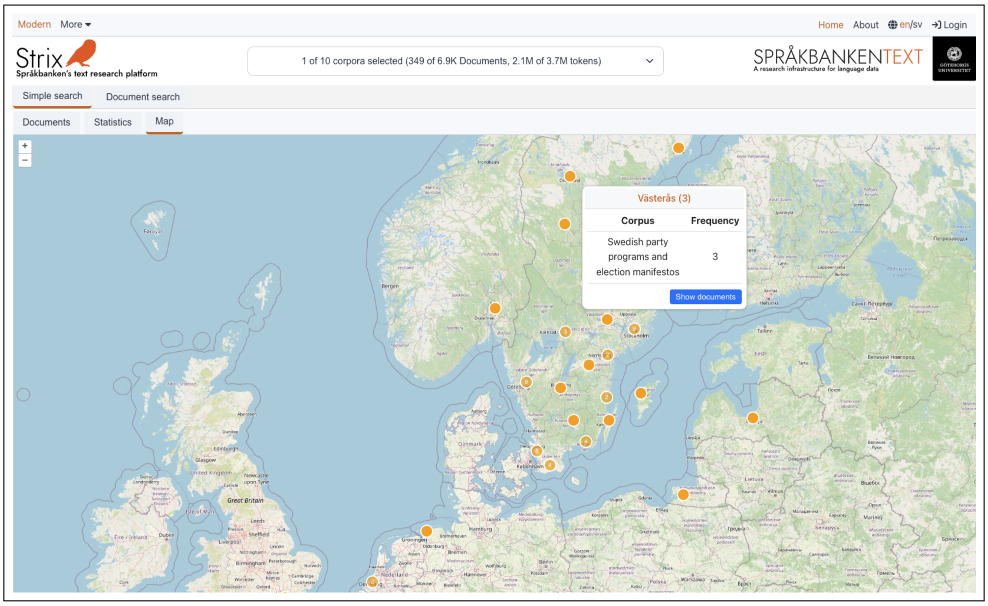
1. **Interactive geo-locations** Each geo-location is represented as a **point** on the map. These points are clickable, allowing users to view detailed information about the number of documents that mention the specific geo-location. If multiple collections are selected, the document counts are separated and displayed in a **tabular format** for clarity (as shown in the figure 10 below).

   Users can click the **"Show documents"** button to open a new tab right beside the **Maps** tab. This tab lists all the documents where the selected geo-location is mentioned.

2. **Handling large datasets** Collections like Wikipedia, which contain hundreds of thousands of geo-locations, are efficiently visualized using **clusters**:

   - **Standalone points**: If a geo-location has no other nearby locations, it is displayed as an individual point.
   - **Clusters**: When multiple geo-locations are close to each other, they are grouped into a cluster. The cluster displays a number indicating how many geo-locations are in that area.

   As users **zoom in**, clusters break apart into individual points, providing a more granular view. Conversely, as users **zoom out**, the points merge back into clusters for a cleaner, high-level overview.


Modern
More
Home
About
en/sv
Login
Strix
Språkbanken's text research platform
1 of 10 corpora selected (349 of 6.9K Documents, 2.1M of 3.7M tokens)
SPRÅKBANKENTEXT
A research infrastructure for language data
Simple search
Document search
Documents
Statistics
Map
+
−

Västerås (3)

Corpus	Frequency
Swedish party programs and election manifestos	3

Figure 10: Popup box for location *Västerås* in the map.

This intuitive and interactive design allows users to explore geographical data effectively, whether they are analyzing a small dataset or navigating through massive collections like Wikipedia. The combination of points, clusters, and detailed document views ensures that users can uncover spatial patterns and relationships with ease.

7 Document View

The **Document view** section in Strix provides users with tools to explore and analyze individual documents in detail. It is divided into two main parts:

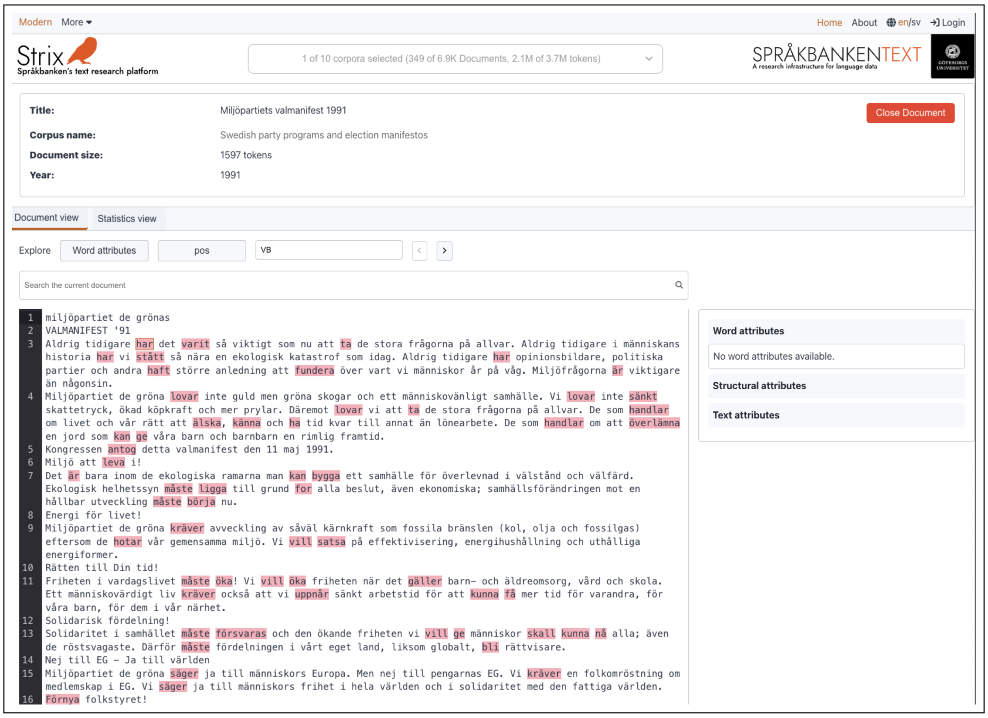
1. **Document reader** This section focuses on visualizing the content of documents, allowing users to explore patterns, trends, and relationships within the text. It provides tools to highlight key information and gain insights into the structure and semantics of the documents.

2. **Document statistics** This section displays the statistics of each word-level metadata attribute for the entire document in a tabular format. Users can analyze word-level details such as part of speech, sentiment, and more, providing a deeper understanding of the document's linguistic and semantic properties.

Explore the subsections to learn more about how each part of the **Document view** works and how it can help you analyze individual documents effectively.

7.1 Document reader

The **Document reader** page in Strix allows users to explore the full content of a selected document in detail. This section focuses on visualizing the document's content, enabling users to identify patterns, trends, and relationships within the text. It provides tools to highlight key information (as shown in figure 11) and gain insights into the document's structure and semantics.



Title: Miljöpartiets valmanifest 1991

Corpus name: Swedish party programs and election manifestos

Document size: 1597 tokens

Year: 1991

pos: VB

1 miljöpartiet de grönas

2 VALMANIFEST '91

3 Aldrig tidigare har det varit så viktigt som nu att ta de stora frågorna på allvar. Aldrig tidigare i människans historia har vi stått så nära en ekologisk katastrof som idag. Aldrig tidigare har opinionsbildare, politiska partier och andra haft större anledning att fundera över vart vi människor år på våg. Miljöfrågorna är viktigare än någonsin.

4 Miljöpartiet de gröna lovar inte guld men gröna skogar och ett människovänligt samhälle. Vi lovar inte sänkt skattetryck, ökad köpkraft och mer prylar. Däremot lovar vi att ta de stora frågorna på allvar. De som handlar om livet och vår rätt att älska, känna och ha tid kvar till annat än lönearbete. De som handlar om att överlämna en jord som kan ge våra barn och barnbarn en rimlig framtid.

5 Kongressen antog detta valmanifest den 11 maj 1991.

6 Miljö att leva i!

7 Det är bara inom de ekologiska ramarna man kan bygga ett samhälle för överlevnad i välstånd och välfärd. Ekologisk helhetssyn måste ligga till grund for alla beslut, även ekonomiska; samhällsförändringen mot en hållbar utveckling måste börja nu.

8 Energi för livet!

9 Miljöpartiet de gröna kräver avveckling av såväl kärnkraft som fossila bränslen (kol, olja och fossilgas) eftersom de hotar vår gemensamma miljö. Vi vill satsa på effektivisering, energihushållning och uthålliga energiformer.

10 Rätten till Din tid!

11 Friheten i vardagslivet måste öka! Vi vill öka friheten när det gäller barn- och äldreomsorg, vård och skola. Ett människovärdigt liv kräver också att vi uppnår sänkt arbetstid för att kunna få mer tid för varandra, för våra barn, för dem i vår närhet.

12 Solidarisk fördelning!

13 Solidaritet i samhället måste försvaras och den ökande friheten vi vill ge människor skall kunna nå alla; även de röstsvagaste. Därför måste fördelningen i vårt eget land, liksom globalt, bli rättvisare.

14 Nej till EG – Ja till världen

15 Miljöpartiet de gröna säger ja till människors Europa. Men nej till pengarnas EG. Vi kräver en folkomröstning om medlemskap i EG. Vi säger ja till människors frihet i hela världen och i solidaritet med den fattiga världen.

16 Förnya folkstyret!

No word attributes available.

Figure 11: Highlighting all the verbs in the document using the annotations selector.

7.1.1 Key features

1. **Full document display** Users can view the entire content of the document, including all text and metadata. This provides a complete overview of the document's structure and content.

2. **Interactive tabs** The document view includes two tabs:

   - **Document tab**: Displays the full content of the document.
   - **Statistics tab**: Provides word-level metadata statistics for the document (accessible via the **Document statistics** section).

   Users can easily switch between these tabs to explore the document's content and metadata.

3. **Annotations and search**

   - **Annotations selector**: Users can navigate through specific annotations or highlights within the document, as shown in figure 11.
   - **Search in document**: A search feature allows users to locate specific words or phrases within the document.

4. **Word metadata** Clicking on a word in the document displays its metadata, such as part of speech, lemma, and other linguistic attributes. This feature is particularly useful for detailed linguistic analysis.

5. **Parallel document mode** For collections that support parallel documents, users can view two documents side by side. This is especially helpful for comparative analysis, such as translations or aligned texts.

6. **Mobile-friendly design** The document view is optimized for mobile devices, with features like collapsible metadata panels and responsive layouts to ensure a seamless experience.

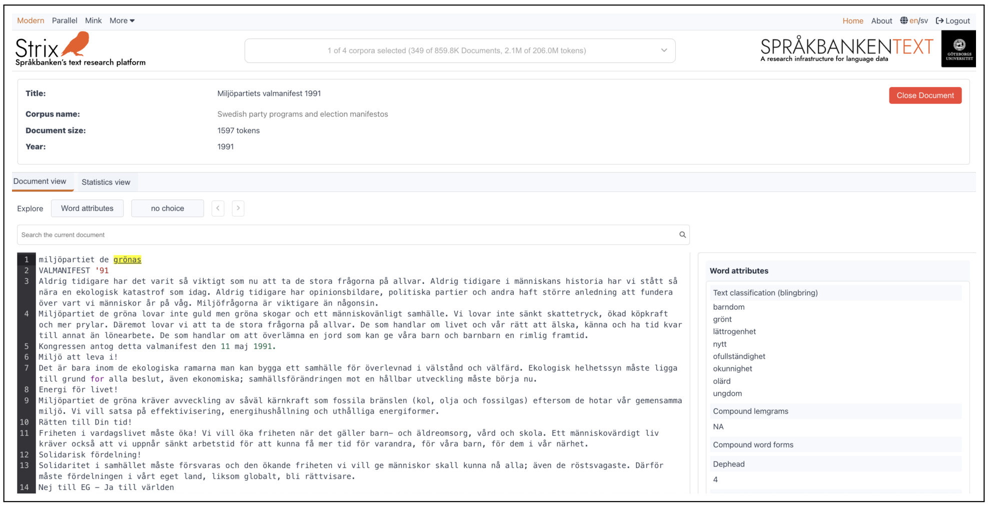
7.1.2 Example view

In figure 12 is an example of the **Document reader** interface, showing the document content, metadata, and interactive features:

This page is designed to give users a comprehensive view of individual documents, making it easier to analyze and extract meaningful insights. Whether you're exploring a single document or comparing parallel texts, the **Document reader** provides all the tools needed for in-depth analysis.

7.2 Document statistics

The **Document statistics** page in Strix provides users with detailed insights into the word-level metadata attributes of a document. This section displays the statistics in a tabular format, allowing users to analyze linguistic and semantic properties of the document.


Modern
Parallel
Mink
More
Home
About
en/sv
Logout
Strix
Språkbanken's text research platform
1 of 4 corpora selected (349 of 859.8K Documents, 2.1M of 206.0M tokens)
SPRÅKBANKENTEXT
A research infrastructure for language data
GÖTEBORGS UNIVERSITET
Close Document
Document view
Statistics view
Explore
Word attributes
no choice
Search the current document

Title: Miljöpartiets valmanifest 1991

Corpus name: Swedish party programs and election manifestos

Document size: 1597 tokens

Year: 1991

1 miljöpartiet de grönas
2 VALMANIFEST '91
3 Aldrig tidigare har det varit så viktigt som nu att ta de stora frågorna på allvar. Aldrig tidigare i människans historia har vi stått så nära en ekologisk katastrof som idag. Aldrig tidigare har opinionsbildare, politiska partier och andra haft större anledning att fundera över vart vi människor år på våg. Miljöfrågorna är viktigare än någonsin.
4 Miljöpartiet de gröna lovar inte guld men gröna skogar och ett människovänligt samhälle. Vi lovar inte sänkt skattetryck, ökad köpkraft och mer prylar. Däremot lovar vi att ta de stora frågorna på allvar. De som handlar om livet och vår rätt att älska, känna och ha tid kvar till annat än lönearbete. De som handlar om att överlämna en jord som kan ge våra barn och barnbarn en rimlig framtid.
5 Kongressen antog detta valmanifest den 11 maj 1991.
6 Miljö att leva i!
7 Det är bara inom de ekologiska ramarna man kan bygga ett samhälle för överlevnad i välstånd och välfärd. Ekologisk helhetssyn måste ligga till grund for alla beslut, även ekonomiska; samhällsförändringen mot en hållbar utveckling måste börja nu.
8 Energi för livet!
9 Miljöpartiet de gröna kräver avveckling av såväl kärnkraft som fossila bränslen (kol, olja och fossilgas) eftersom de hotar vår gemensamma miljö. Vi vill satsa på effektivisering, energihushållning och uthålliga energiformer.
10 Rätten till Din tid!
11 Friheten i vardagslivet måste öka! Vi vill öka friheten när det gäller barn- och äldreomsorg, vård och skola. Ett människovärdigt liv kräver också att vi uppnår sänkt arbetstid för att kunna få mer tid för varandra, för våra barn, för dem i vår närhet.
12 Solidarisk fördelning!
13 Solidaritet i samhället måste försvaras och den ökande friheten vi vill ge människor skall kunna nå alla; även de röstsvagaste. Därför måste fördelningen i vårt eget land, liksom globalt, bli rättvisare.
14 Nej till EG – Ja till världen

Word attributes

Text classification (blingbring)
barndom
grönt
lättrogenhet
nytt
ofullständighet
okunnighet
olärd
ungdom

Compound lemgrams
NA

Compound word forms

Dephead
4

Figure 12: Document reader view in Strix.

7.2.1 Key features

1. **Word-level metadata attributes** Users can explore various word-level metadata attributes, such as part of speech, lemma, sentiment, and more. These attributes provide a deeper understanding of the document's linguistic structure.

2. **Dynamic attribute selection**

    - A list of available word-level metadata attributes is displayed on the left side of the interface as shown in figure 13.
    - Users can select an attribute to view its statistics in the table.
    - By default, the first attribute in the list is selected when the page loads.

3. **Tabular statistics** The statistics for the selected metadata attribute are displayed in a table on the right side in figure 13. The table includes:

    - **Attribute elements**: The unique values or elements of the selected metadata attribute (e.g., specific parts of speech or lemmas).
    - **Frequency**: The number of occurrences of each element in the document.

4. **Filtering and Pagination**

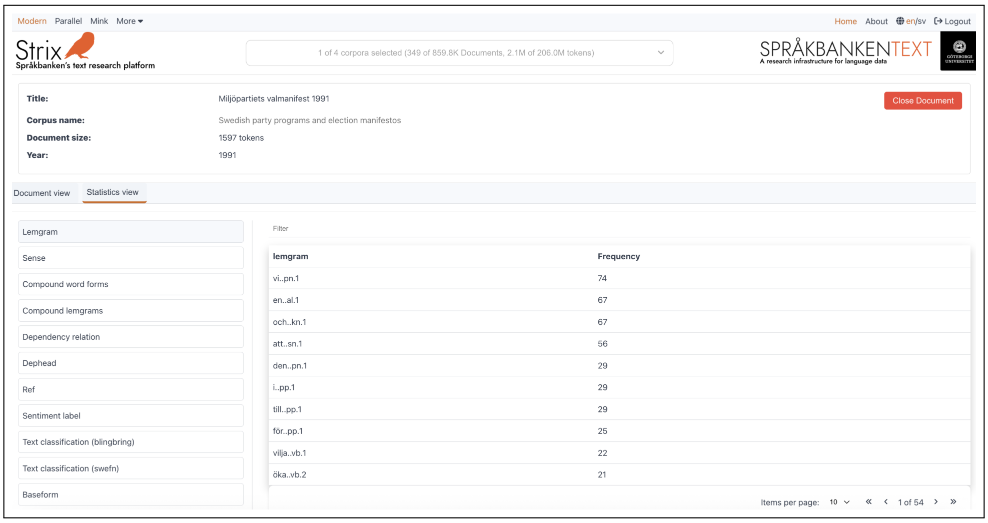
    - Users can filter the table by entering a keyword in the search box to narrow down the results.
    - Pagination controls allow users to navigate through large datasets, with options to adjust the number of items displayed per page.

5. **Interactive design**

    - The interface is responsive and optimized for both desktop and mobile devices.
    - On mobile, a dropdown menu is used for selecting metadata attributes, ensuring a seamless experience.

7.2.2 Example view

Below is an example of the **Document statistics** interface, showing the metadata attributes and their statistics:



1 of 4 corpora selected (349 of 859.8K Documents, 2.1M of 206.0M tokens)

Title: Miljöpartiets valmanifest 1991

Corpus name: Swedish party programs and election manifestos

Document size: 1597 tokens

Year: 1991

lemgram	Frequency
vi..pn.1	74
en..al.1	67
och..kn.1	67
att..sn.1	56
den..pn.1	29
i..pp.1	29
till..pp.1	29
för..pp.1	25
vilja..vb.1	22
öka..vb.2	21

Items per page: 10 « ‹ 1 of 54 › »

Figure 13: Document statistics interface displaying statistics for the word attribute *lemgram*.

7.2.3 How it works

1. **Selecting a metadata attribute**
   - On desktop, users can click on an attribute from the list on the left.
   - On mobile, users can select an attribute from the dropdown menu.
   - The table updates dynamically to display the statistics for the selected attribute.
2. **Filtering the data**
   - Enter a keyword in the search box to filter the table and display only the relevant elements.
   - The filtering is case-insensitive and works across all elements in the table.
3. **Navigating the table**
   - Use the pagination controls to navigate through the table.
   - Adjust the number of items displayed per page using the dropdown menu.

This page is designed to provide users with a detailed and interactive way to analyze the word-level metadata of a document, making it easier to uncover linguistic patterns and insights.

8 Related Documents

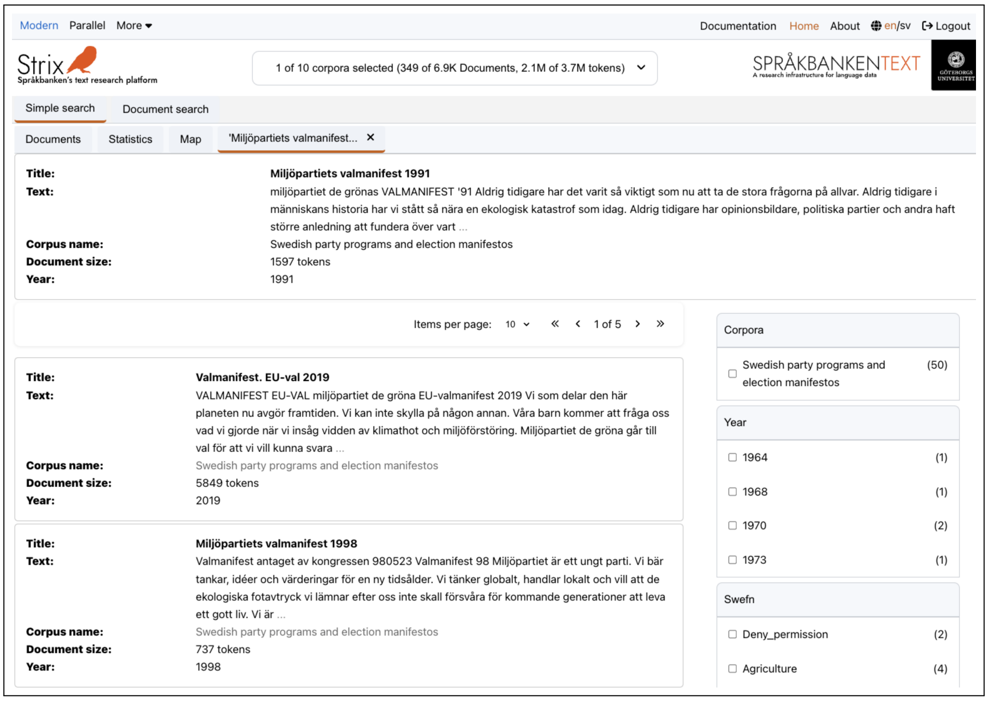
The **Related documents** section in Strix allows users to explore documents that are semantically similar to a selected document. This feature is designed to help users uncover connections, patterns, and relationships across the dataset, enabling deeper analysis and discovery.

8.1 Key features

1. **Top related documents**
   - Strix retrieves a ranked list of related documents based on semantic similarity.
   - Users can view the **top 10, 20, 25, or 50 related documents**, depending on their selection.
   - Each related document is displayed with key details (see details below).
2. **Interactive document preview**
   - Each related document includes the following key details:
     - **Title**: The title of the document.
     - **Snippet**: A short excerpt or highlighted text from the document to provide context.
     - **Corpus name**: The collection to which the document belongs.
     - **Document size**: The number of tokens (words or word-like units) in the document.
     - **Year**: The year the document was created (if available).
     - **Source link**: A clickable link to the document's source (if provided in the metadata).
3. **Graph visualization**
   - For certain modes, users can visualize the relationships between the selected document and its related documents using a **graph view**.
   - The graph displays nodes (documents) and edges (connections), with the size and color of nodes representing their similarity scores and metadata attributes.
   - Users can toggle between the **graph view** and the **document list view** for flexibility.
4. **Filtering and pagination**
   - Users can filter related documents by metadata attributes such as **corpus**, **year**, **sweFN**, and **blingbring**.
   - Pagination controls allow users to navigate through the list of related documents, with options to adjust the number of items displayed per page.
5. **Mobile-friendly design**
   - The interface is optimized for mobile devices, ensuring a seamless experience with responsive layouts and collapsible controls.

8.2 Example view

Below is an example of the **Related documents** interface, showing the list of semantically similar documents and the graph visualization:



Title: **Miljöpartiets valmanifest 1991**
Text: miljöpartiet de grönas VALMANIFEST '91 Aldrig tidigare har det varit så viktigt som nu att ta de stora frågorna på allvar. Aldrig tidigare i människans historia har vi stått så nära en ekologisk katastrof som idag. Aldrig tidigare har opinionsbildare, politiska partier och andra haft större anledning att fundera över vart ...
Corpus name: Swedish party programs and election manifestos
Document size: 1597 tokens
Year: 1991

Title: **Valmanifest. EU-val 2019**
Text: VALMANIFEST EU-VAL miljöpartiet de gröna EU-valmanifest 2019 Vi som delar den här planeten nu avgör framtiden. Vi kan inte skylla på någon annan. Våra barn kommer att fråga oss vad vi gjorde när vi insåg vidden av klimathot och miljöförstöring. Miljöpartiet de gröna går till val för att vi vill kunna svara ...
Corpus name: Swedish party programs and election manifestos
Document size: 5849 tokens
Year: 2019

Title: **Miljöpartiets valmanifest 1998**
Text: Valmanifest antaget av kongressen 980523 Valmanifest 98 Miljöpartiet är ett ungt parti. Vi bär tankar, idéer och värderingar för en ny tidsålder. Vi tänker globalt, handlar lokalt och vill att de ekologiska fotavtryck vi lämnar efter oss inte skall försvåra för kommande generationer att leva ett gott liv. Vi är ...
Corpus name: Swedish party programs and election manifestos
Document size: 737 tokens
Year: 1998

Figure 14: Related documents in Strix.

8.3 How it works

1. **Viewing related documents**
   - When a user selects a document, Strix retrieves the top related documents based on their semantic similarity.
   - Users can explore these documents in a **list view** or switch to the **graph view** for a visual representation of relationships (if available).
2. **Exploring the graph view**
   - The graph displays related documents as nodes, with edges representing their connections to the selected document.
   - Users can interact with the graph by zooming in/out or clicking on nodes to view more details.
3. **Filtering and navigation**
   - Use the filtering options to refine the list of related documents based on specific metadata attributes.
   - Navigate through the list using pagination controls and adjust the number of items displayed per page.

This feature is designed to provide users with an intuitive and interactive way to explore related content, making it easier to uncover meaningful relationships and gain deeper insights into the dataset.

9 Login Access

Strix provides access to advanced text analysis tools and datasets. Some datasets in Strix are protected and require login access. Below are the details on how to gain access to Strix.

9.1 Who can access Strix?

1. **Academic users**: If you are affiliated with a university or academic institution, you can log in using your institutional credentials through the eduGAIN network. This includes most researchers, faculty, and students.

2. **Other users**: If you are not affiliated with an academic institution, you can create an account through eduID. eduID is a secure identity provider that connects to the eduGAIN network, enabling access to Strix.

9.2 Steps to gain access

9.2.1 For academic users:

1. Visit the Strix login page.

2. Select your institution from the list of eduGAIN-supported organizations.

3. Log in using your institutional credentials.

9.2.2 For non-academic users:

1. Create an account at eduID.

2. Verify your identity as part of the registration process.

3. Once your eduID account is active, use it to log in to Strix.

9.3 Why is login required?

Login access is required to protect sensitive datasets. By restricting access to verified users, Strix ensures the integrity of its datasets and tools while providing a secure environment for research and analysis.

9.4 Troubleshooting login issues

If you encounter any issues while logging in:

- Ensure that your institution is part of the eduGAIN network. You can check the list of supported organizations on the eduGAIN website.

- For eduID users, ensure that your account is verified and active.

- If problems persist, contact the Strix support team.

If you have further questions about login access or need assistance, feel free to reach out to the Strix support team at sb-info@svenska.gu.se.