



Artificial intelligence at IP7

Table of contents

Use of AI	2
General information	2
Folder assignment	3
Classifier	3
Training statistics.....	4
Limitations	5
Automatic classification	5
Limitations.....	6

Use of AI

General information

Too many patents and too little time

In cooperation with Averbis, a leading provider in the field of artificial intelligence, we have developed a concept that simplifies specific processes within our software and increases efficiency. By using natural language models, Averbis analyzes large volumes of patent documents to classify them according to your individual requirements and guidelines using machine learning.

(Averbis: <https://averbis.com/de/patent-kategorisierung-und-monitoring/>)

The AI can be used in various areas within the IP7 Compass:

- ❖ Folder assignment or automatic classification of patents
- ❖ Automatic sorting out of "irrelevant" hits (coming soon!)
- ❖ Sorting of result lists according to relevance (also available soon!)

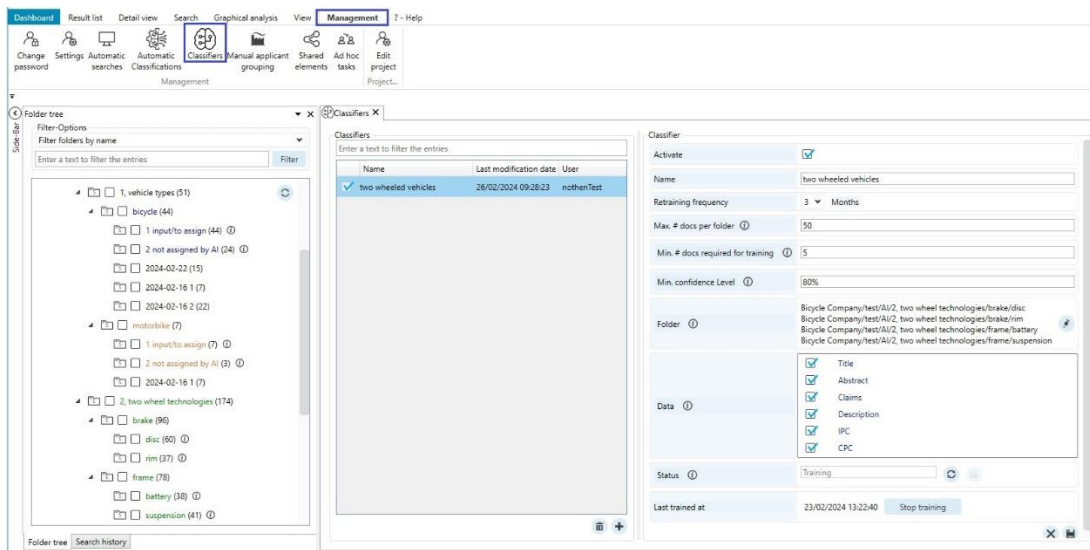
The main characteristic of AI is that it is trained using predefined data and "learns" in the process. The aim is to continuously improve the AI through constant human adjustments and repeated training. The quality of the artificial intelligence is therefore always dependent on the data provided and human support or correction.

We are currently in the pilot phase. It is possible that individual functions, interfaces and the scope of services will change in the future. AI is not integrated into Compass as standard, but there is an additional module that can be activated on request by IP7. There are extra costs for using this technology. We would be happy to provide you with an individual quote for this.

Folder assignment

Folder structures are created in the IP7 Compass to classify patents according to individual criteria. For example, a technology tree can be created that allows patents to be assigned to corresponding technologies, including business areas or files. The "Automatic classification" function allows artificial intelligence to perform this assignment task independently.

Classifier



Before artificial intelligence is able to classify patents into folders, it needs to learn from existing classifications. This requires an existing folder structure with previously correctly classified patents. A "classifier" can then be created. During this process, various settings are defined, many of which have a direct influence on the training process:

- **Activate:** This is where you decide whether automatic training should take place regularly or whether the classifier can be used.
- **Name**
- **Training frequency:** Determines the frequency of automatic training of the classifier. As the contents of the folders change over time, it makes sense to have the classifier trained regularly.
- **Maximum number:** Defines the maximum number of documents that the AI may use per folder for training purposes. Too few documents per folder can reduce the training quality, while too many can significantly increase the training duration. A constant number of documents per folder is ideal.

Example: Suppose you want to classify a large number of folders automatically. Most of the folders contain around 50 documents. However, a few folders contain more than 500 documents. In this case, the value for the maximum number should be set to 50 documents in order to achieve balanced training and therefore better results. The upper

limit is 1,000 documents per folder. If there are more documents in a folder, 1,000 documents are randomly selected from this folder.

- **Minimum number:** Specifies the minimum number of documents that must be present in a folder for them to be included in the training.

Example: the AI cannot gain any good insights from a folder with only 3 documents. The minimum value is 10 documents. Folders with a lower number are excluded from the training.

- **Minimum confidence level:** The "Confidence Level" indicates as a percentage how confident the AI is in an assignment. Everything below the minimum value is moved to the "unclassifiable" folder. This is defined later in the automatic classification.
- **Folder:** Determines which folders are used for training. The "automatic classification" will later make the assignments to these folders. Up to 500 folders can be selected.
- **Data:** Determines which data the AI receives for the training (title, summary, claims, description, IPC, CPC). Selecting "Description" provides the AI with significantly more training data, which will have a corresponding effect on the duration of the training.
- **Status:** Shows the current status of the classifier:

"Training required": Training is required after the classifier has been created.

"Training": The classifier is in the training process. During this time, no automatic classifications using this classifier can be carried out.

"Ready": The classifier is ready for automatic classification, which also means that the last training has been successfully completed.

"Error": An error has occurred. Successful training is a prerequisite for use in automatic classification.

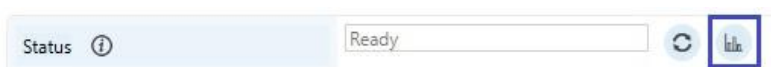
- **Last training:** This shows when the last successful training session was carried out. The current status can be called up using the "Refresh" button:



As soon as the "Classifier" has finished training ("Ready" status), it can be used in an automatic classification.

Training statistics

An analysis of the training run can be triggered by the following button:



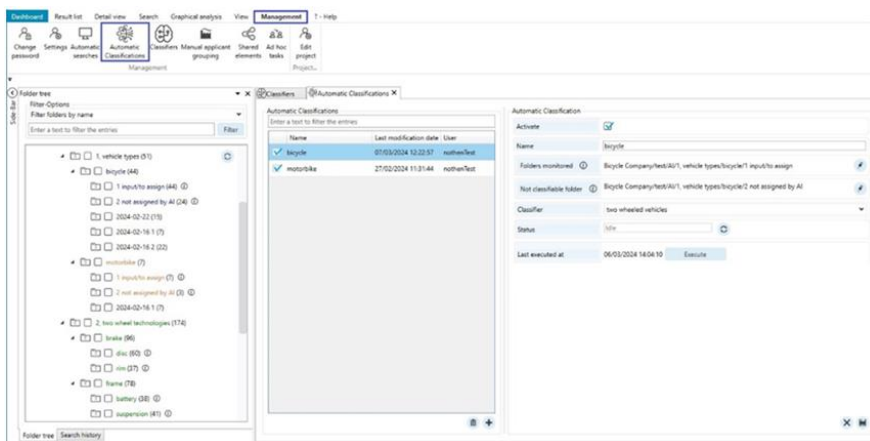
Further information on the terms Precision, F1Score and Recall:

https://en.wikipedia.org/wiki/Precision_and_recall

Limitations

A maximum of 10,000 patents from all folders included in the classifier are used for training. If the total number of patents in the folders exceeds this number, they are reduced proportionally. If this causes the number of patents in a folder to fall below the specified minimum quantity, the number of selected patents in this folder is increased again up to the minimum limit. This means that the actual upper limit can easily exceed 10,000.

Automatic classification



The following setting options are available when setting up automatic classification:

- **Activate:** This option determines whether the classification is carried out automatically. This option can be used to quickly stop an active automatic classification in the event of problems.
- **Name**
- **Monitored folders:** Determines the "input folders" whose contents are classified by the AI. This includes all current and future patents in these folders. The inbox folders should not be confused with the classifier folders: The folders into which the AI assigns the patents are defined in the Classifier.
- **Folder for "not classifiable":** All patents that cannot be classified by the AI are filed here.
- **Classifier:** Selection of the previously created classifier.
- **Status:**

"idle" (inactive) - The automatic classification is not running at the time.
"running" (active) - The automatic classification is currently running.
- **Last execution:** The last execution of the automatic classification.

Multiple automatic classifications:

A classifier can theoretically be used for several automatic classifications.

Example:

Several vehicle types exist that are to be monitored and then assigned to a technology tree:

1 "Vehicle types" -> "Bicycle" and "Motorcycle"

There is a folder structure or a technology tree that is to include all types of two-wheeled vehicles:

2 "Technologies for two-wheelers"

- ➔ However, the results that cannot be clearly assigned by the AI must be saved separately. For this reason, a separate automatic classification is created for "bicycle" and "motorcycle". However, the classifier only needs to be created once for this case.

Limitations

Up to 5,000 hits or patents can be classified in one run.

If the number of patents exceeds this limit, automatic classification is not carried out and is automatically deactivated. As soon as the run is started manually, a corresponding warning message appears. If the run is started anyway, the classification is limited to a maximum of 5,000 patents.

Patents already classified in a run do not have to be removed from the input folder. They are identified as already classified in subsequent runs and are not processed again. In order to have one or more patents classified again, it is necessary to first remove them from the folder and then add them again. With this new assignment, they are no longer considered as already classified.