

Disco: Discover Your Processes

Christian W. Günther and Anne Rozinat

Fluxicon

Bomanshof 259, 5611 NS, Eindhoven, The Netherlands.

{christian,anne}@fluxicon.com

Abstract. Disco is a complete process mining toolkit from Fluxicon that makes process mining fast, easy, and simply fun.

1 Why Disco?

As former process mining researchers, we started Fluxicon in 2009 to build professional tools that help organizations to regain control over their processes.

Our first product Nitro addressed the pain of getting the original process data from IT systems into a format that can be used for process mining. Today, Nitro is used all over the world by practitioners and researchers to convert raw data into event logs that can be analyzed with the leading academic process mining toolkit ProM.

While ProM is great and immensely powerful, we realized through our own process mining consulting projects, and through many conversations with practitioners, that process analysts in practice need a tool that—above all—makes process mining easy and fast. And this is what Disco is all about.

2 Tour

The following tour gives you an overview about the main functionality of Disco.

2.1 Import

Every process mining project starts with the data that should be analyzed. Disco has been designed to make the data import really easy by automatically detecting timestamps, remembering your configuration settings, and by loading data sets with high speed.

One simply opens a CSV or Excel file and configures which columns hold the case ID, timestamps, activity names, which other attributes should be included in the analysis, and the import can be started. Data sets are imported in a read-only mode, so the original files cannot be modified (which is important, e.g., for auditors).

Disco is also fully compatible with the academic toolsets ProM 5 and ProM 6. By importing and exporting the event log standard formats MXML and XES, advanced users can seamlessly move back and forth between Disco and ProM if they want to benefit from the new research technologies developed in academia.

Disco also features a short-cut import and data exchange for previously imported data sets with up to 200x speed-up for very large data sets through the native FXL Disco log file format.

2.2 Automated Process Discovery

The core functionality of process mining is the automated discovery of process maps by interpreting the sequences of activities in the imported log file. After one presses the Start import button the user is taken right into the Map view, where she can quickly and objectively see how the process has been actually performed.

Disco uses an intuitively understandable and 100% truthful process map visualization. The thickness of paths and coloring of activities show the main paths of the process flows, and wasteful rework loops are quickly discovered.

The Disco miner is based on the Fuzzy miner, but has been further developed in many ways. The Fuzzy Miner was the first mining algorithm to introduce the “map metaphor” to process mining, including advanced features like seamless process simplification and highlighting of frequent activities and paths. For Disco, we have used the approach of the Fuzzy Miner and combined it with experience from our own practice and user testing.

The result is a mining algorithm that, while providing reliable and trustworthy results for data sets of arbitrary complexity, can be operated and understood efficiently by domain experts with no prior experience in process mining. Although the Disco miner is based on the framework of the Fuzzy Miner, we have developed a completely new set of process metrics and modeling strategies, effectively making the Disco miner a next-generation Fuzzy Miner.

Our design priorities are what sets the Disco miner apart from other solutions:

1. *Usability*: Our goal was to have a miner that can be operated and understood by domain experts, with an adequate learning curve to also accommodate process mining experts. We also have put great effort into making our visualizations information-dense, while avoiding information overload. For Disco, we have used state-of-the-art UX and visualization research, user testing, and lots of development time to make sure our models are nice to read and quick to understand.
2. *Fidelity*: Creating a truthful model from a simple, well-structured process model is easy. When faced with complex data, though, most commercial approaches resort to drastically limiting the data used (only using the mainstream variants) to keep model complexity in check. We wanted a miner that can intelligently extract the most important parts of the process from the full set of data, and create a useful process model from data of arbitrary complexity.
3. *Performance*: Almost all process mining tools want to be used in a procedural fashion: You give them the data, and some parameters, they create a process model, done. We see process mining as an explorative and highly interactive task, where the domain expert learns to understand the data by looking at the process from multiple perspectives in quick succession. For this approach to work, we need our miner to work very fast.

The Disco miner is considerably faster than any other approaches we are aware of, while delivering superior model quality. We think there is inherent value in having a good approximation of complex behavior in a few seconds, versus a perfect model in three hours (which is what you get with, e.g., genetic approaches). By intensively optimizing the whole stack, down from the log storage layer up to the graph visualization,

we have created a miner that fosters truly interactive usage which, ultimately, leads to better and more meaningful analysis results.

2.3 Process Statistics

Next to the process maps one can also inspect statistics about the process. For this, one simply changes to the Statistics tab in the toolbar. The user will get overview information about the number of cases and events in the data set, the time frame covered, and performance charts like, for example, about the case duration.

Further statistics views provide frequency and performance information for all activities and resources in the process. Furthermore, there are statistics for any additional data attribute column that was included in the data set. These additional data attributes are usually very important for the process analysis, because they hold relevant context information such as:

- Which *product* a service call was about,
- Which type of *category* a change request in an IT Service process falls in,
- The *channel* through which a lead in a sales process came in,
- *Domain-specific characteristics* such as warranty vs. out-of-warranty repairs in a service process,
- By which *department* the activity was handled,
- In which *country* the process was performed,
- The *value* of an order, which is relevant for many purchasing processes, because depending on the amount of money that is involved different anti-fraud rules will apply, etc.

In our projects, we often get data sets with up to 40 or 60 additional data attributes that are relevant and can be used in the analysis. Disco shows the users these attribute statistics, but also lets them use them to drill down and focus their analysis, and to split out and compare processes with respect to these categories.

2.4 Variants and Individual Cases

The third data set view is the Cases tab. While the Map view gives an understanding about the process flows, and the Statistics view provides detailed performance metrics about the process, the Cases view actually goes down to the individual case level and shows the raw data.

To be able to inspect individual cases is important, because one will need to verify the findings and see concrete examples particularly for “strange” behavior that will most likely be discovered in the process analysis. Almost always users find things that are hard to believe until they have drilled down to an individual example case, noted down the case number, and verified that this is indeed what happened in the operational system.

Furthermore, looking at individual cases with their history and all their attributes can give additional context (like a comment field) that sometimes explains why something happened. Finally, being able to drill down to individual cases is important to be able to

act on the analysis. For example, if one has found deviations from the described process, or violations of an important business rule, one may want to get a list of these cases and talk to the people involved in them to provide additional training.

In addition to a complete list of all cases in the data set, the user also gets direct access to the variants in the process. Variants are an integral part of the process analysis. In Disco, a variant is a specific sequence of activities. It can be seen as one path from the beginning to the very end of the process. In the process map, an overview of the process flow between activities is shown for all cases together. A variant is then one “run” through this process from the start to the stop symbol, where also loops are unfolded. Usually, a large portion of cases in the data set are following just a few variants, and it is useful to know which are the most frequent ones.

Furthermore, a live full text search across case names and all activity, resource, and data columns lets the user find specific cases based on the words or word fragments she is looking for.

2.5 Filtering

Disco offers powerful, non-destructive filtering capabilities for explorative drill-down, and for focusing the analysis. These filters are quickly accessible from any view and easy to configure.

In total, there are six powerful filter types available in Disco, and they can be combined and stacked in any order:

- The *Timeframe filter* with intuitive calendar controls to select cases and events based on a time window. It can be used, for example, to compare the processes before and after a process change.
- The *Variation filter* that allows one to focus the analysis on either the mainstream behavior or precisely the exceptional cases by making use of the variants from the Cases view.
- The *Performance filter* to focus on cases based on a variety of different performance metrics like, for example, the case duration.
- The *Endpoints filter* to select cases based on their start and end activities. For example, one can filter incomplete cases, or trim cases to cut out a part of the process.
- The *Attribute filter* to focus on (or exclude) certain activities, resources or process categories based on data attributes.
- The *Follower filter* for powerful process pattern-oriented filtering, including a 4-Eyes filter option that can be used to check for segregation of duty violations.

Together with the three analysis views, these filtering capabilities enable Disco users to quickly and interactively explore their process into multiple directions, and to answer concrete questions about the process. Because filtering, and Disco in general, are so fast, one can also hold interactive process workshops, where the analyst and a group of other process stakeholders get together to do an As-Is analysis and generate process improvement ideas along the way.

2.6 Performance Highlighting

In addition to the frequency-based process map, one can also analyze the time that is spent in the process. The average durations of the activities and the inactive (waiting) times between activities are automatically extracted from the timestamps in the data set and visually projected onto the process map.

An alternative Total durations performance highlighting option shows these high-impact areas at one glance by summing up the durations for each activity and path for the complete data set.

2.7 Animation

Animation is a way to visualize the process flow over time right in the discovered process map (a bit like showing a “movie” of the process). Animation should not be confused with simulation. Rather than simulating, the real events from the log are replayed in the discovered process map as they took place.

Animation can be very useful to communicate analysis results to process managers or other people who are no process analysis experts. By showing how the cases in the data set move through the process (at their relative, actual speed), the process is literally “brought to life”.

2.8 Project Management

One of the advantages of Disco is that it supports project work through the management of multiple data sets in one project view. In a typical process mining project, one will import log files in different ways, filter them, and make copies to save intermediate results. This results in many different versions and views of the data sets and can easily get out of hand.

The project view in Disco is there to help the users keep an overview. It keeps all their work in one place and lets them make notes about what they found out, or what they still want to check. Complete projects can be exported and shared with other people who can start right where they left off.

Disco features a sandbox project that we prepared for new users to get started quickly after the installation of Disco.

3 Links

A 6-min screencast has been recorded for this demo. You can watch this screencast in two parts, Part I at <http://screenr.com/F1n8>, and Part II at <http://screenr.com/q1n8>.

Furthermore, you can view the Disco product page and download a free demo version at <http://fluxicon.com/disco/>. You can also read a tour including screenshots and examples in our launch blog post here: <http://fluxicon.com/blog/2012/05/say-hello-to-disco/>.

Note that we provide free academic licenses for Disco in our Academic Initiative for Process Mining Research and Education (see <http://fluxicon.com/academic/>).