

THE EXTRACT TOOL

Manually extracting the data you need from a PDF of free text, tables, graphs, figures and a plethora of layouts requires major effort from highly skilled manual labor. Iris.ai's Extract tool fetches all the key data from these documents into a tabular, machine readable, systematic format. A full month of work can be done in minutes.

HOW IT WORKS

The PDF to be extracted is sent to the Iris.ai system. This PDF can be a patent, a clinical trial report, a research paper or any other relevant type of scientific content. It can be one simple document at a time, or hundreds or thousands of them in a batch. The Iris.ai engine extracts the text and identifies all the domain specific entities, then locates the tables and extracts the data from rows and columns, and links the data between the text and table. Graphs, figures and other elements go through the same process. Then the engine populates a pre-defined output in a machine readable format; an excel sheet, an integrated lab tool, a database or anywhere else your researchers require.

CONTACT

Amin Mehdipoor
amin@iris.ai

PROVEN PERFORMANCE

90%

The Extract tool has a 90% precision and recall, higher than the average human performance.



The research team can now focus on far more exciting and value creating tasks.



The average processing time of a PDF is less than 2 minutes, where a human would use half a day.



Training requires minimal human effort, only a set of documents in the domain and optionally an ontology of entities of interest.

WHY IRIS.AI?

Iris.ai has spent the last 5 years building an award winning AI engine for scientific text understanding. Our algorithms for text similarity, tabular data extraction, domain specific entity representation learning and entity disambiguation and linking measure up to the best in the world. On top of that

our machine builds a comprehensive knowledge graph containing all entities and their linkages to allow humans to learn from it, use it and also give feedback to the system. Applying these on scientific and technical text is a complicated challenge few others can achieve.

