

## Chapter 2: RapidMiner



# 1. Presentation of RapidMiner and Natural Language Processing

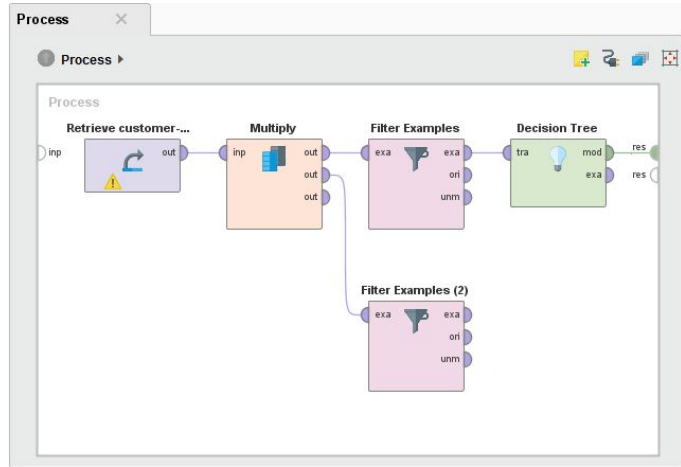


**RapidMiner** is a data science software that will help us to analyse data without writing any lines of code.

So instead of something that might look like this.



```
31 def __init__(self, job_dir):
32     self.file = None
33     self.fingerprints = set()
34     self.logdups = True
35     self.debug = debug
36     self.logger = logging.getLogger(__name__)
37     if path:
38         self.file = open(os.path.join(path, "requests.txt"),
39                         "w")
40         self.file.seek(0)
41         self.fingerprints.update(s.request() for s in self.requests)
42
43 @classmethod
44 def from_settings(cls, settings):
45     debug = settings.getbool("debug", False)
46     return cls(job_dir(settings), debug)
47
48 def request_seen(self, request):
49     fp = self.request_fingerprint(request)
50     if fp in self.fingerprints:
51         return True
52     self.fingerprints.add(fp)
53     if self.file:
54         self.file.write(fp + os.linesep)
55
56 def request_fingerprint(self, request):
57     return request_fingerprint(request)
```



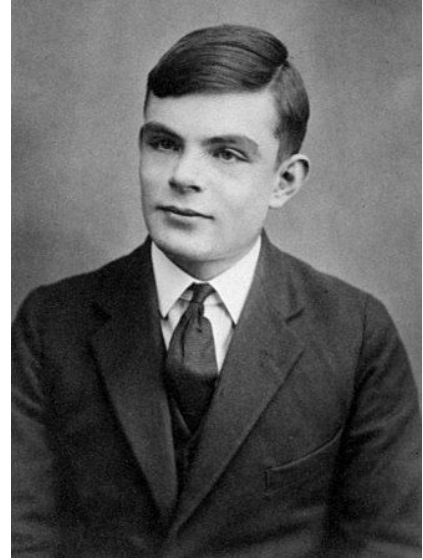
Rather, we will have something like that.

In the previous topic we were talking about AI and so RapidMiner gives us the possibility to import our datas and to analyse them in multiple ways using **Natural Language Processing (NLP)**.

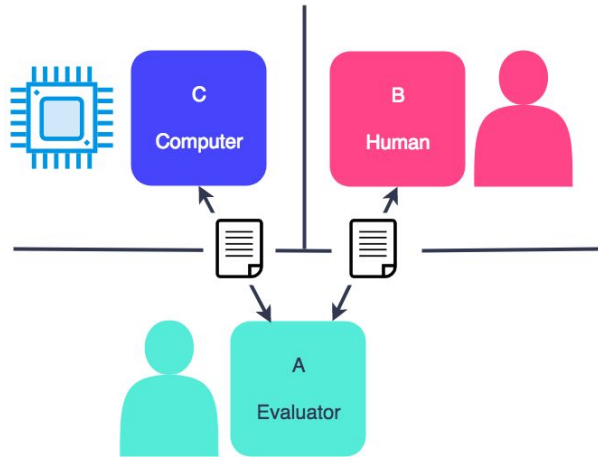
But what is  
**NLP** ?



**Natural language processing** has its roots in the 1950s. Already in 1950, **Alan Turing** considered to be the father of theoretical computer science and artificial intelligence proposed what is now called the **Turing test** as a criterion of intelligence.



*Alan Turing at the age of 16*



The **Turing test** is a test of a machine's ability to exhibit intelligent behavior equivalent to, or indistinguishable from, that of a human.



That's a lot of information isn't it ?

So to make it simple, **natural language processing** is the capacity of an AI to understand the contents of documents including the contextual nuances of the language within them.

