Electronic discovery is the electronic aspect of identifying, collecting and producing electronically stored information (ESI) in response to a request for production in a law suit or investigation. Researchers would typically have to review documents manually. Because it is impossible to review all documents in the dataset by hand, the amount of results is often limited by keyword searches.

The use of electronic discovery allows for the identification of groups or clusters of related documents was carried out using a combination of text analysis and network analysis. The key assumption here is that documents can be linked when they share similar content in their abstracts. TFIDF (term frequency – inverse document frequency) models are a common technique in text analysis to assign numerical vectors to documents (here, their abstracts), which can then be used to calculate pairwise similarities between all documents in the dataset.

So what is the business case for electronic discovery usage? It reduces timelines, costs, and reliance on IT for legal holds and data preservation. In a complex lawsuit, it is estimated that nearly 75% of overall costs is in document research. The use of electronic discovery greatly reduces this cost and timeline.

<https://cdslegal.com/knowledge/the-basics-what-is-e-discovery/>

<https://beta.vu.nl/nl/Images/stageverslag-jeukendrup_tcm235-449790.pdf>