



*VESTIGO*

Teemu Vesterinen

Document Similarity Search

# Table of contents



1. Why Document Similarity Search (DSS) ?
2. When to use Document Similarity Search (DSS) ?
3. Reference document
4. Data set search
5. Similarity search
6. Search results

# Why Document Similarity Search?



1. To supplement manual searches
2. Fast and comprehensive results
3. Searches across classification boundaries
4. Extensive document similarity calculations
  - Using original full text parts from documents, not just keywords
  - Multiple calculations to get better approximations

# When to use Document Similarity Search?



## 1. Normal prior art searches

- To supplement manual search

## 2. Competitor monitoring

## 3. Licensing & litigations

- Searches for patent infringements
  - Which companies most likely a claim under DSS test
  - Which companies to approach
- Defending against patent infringements
  - The validity of offensive patent(s)

## Reference document



- Delivered by the client
- Should contain at least claims and description
- The more comprehensive a document, the better the results
- The publication number is enough

# Data set search



- Data sets provided by:
  - Customer
    - Keywords for Espacenet search

or

  - Vestigo
    - EPO Espacenet

## Data set search

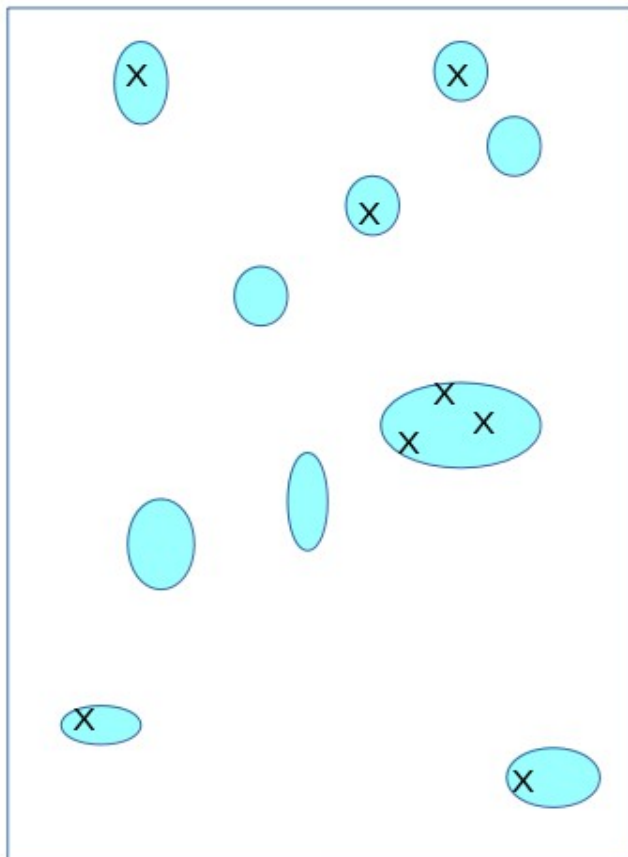


- Data sets will be downloaded from the EPO OPS service
  - Search results will be slightly different from manual Espacenet search results

# Similarity search – Manual vs. DSS



● = Search result (multiple patents)  
X = Prior-art from manual search



## Manual search:

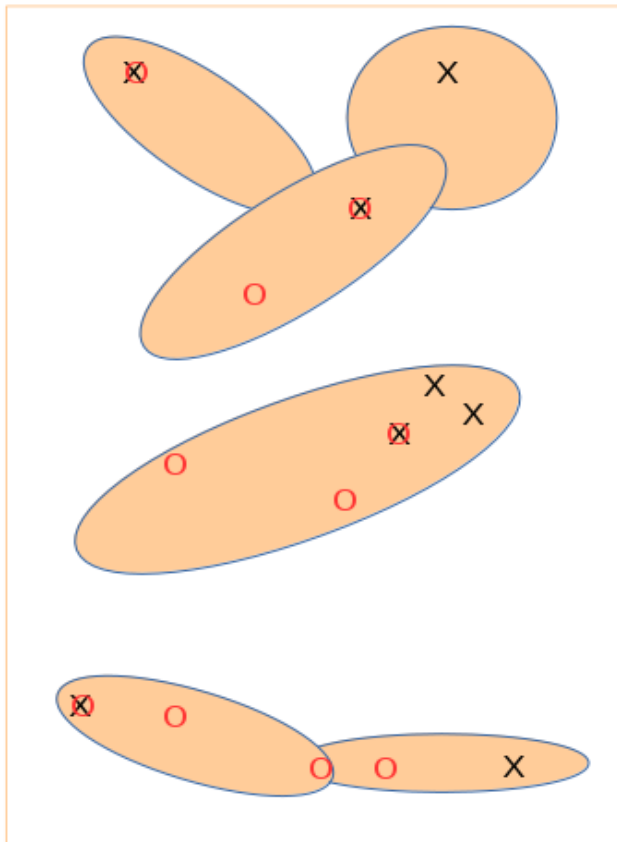
- Multiple keywords to reduce search results
- Small search result data sets
- No automatic prioritizing
- Read all the search result documents



# Similarity search - Manual vs. DSS



- = DSS search result (multiple patents)
- X = Prior-art from manual search
- = Prior-art from DSS search



## DSS search with full benefits:

- Less limiting keywords
- Bigger data sets
- Full document calculations
- Document prioritizing
- Enables selective reading from the suggested best documents - ○

## Search results



- Different methods to calculate similarities between a reference document and the data set documents
- All the methods presented separately
- Different similarity checks to provide the final star ranking for the documents (one star for every check)
- All the results represented such that the most similar document is presented first.

## Similarity search results



- Processing data sets comprising less than 10 000 documents takes one business day.
- The results are provided in PDF - format and will be delivered via email



**Thank You for the Attention**