Search Matters 2016

The EPO pre-search framework

Johan Bengtsson

Examiner, Directorate 1507

9-10 June 2016

Agenda

1. Background of pre-search
2. The pre-search algorithms
3. Input for pre-search algorithms
4. The bibliographic search
5. The class-based search
6. The term based search
7. Presentation of the results from pre-search
8. Evaluation of the pre-search algorithms
9. Working with results of an automated search
10. Future of pre-search
11. Conclusions

Background of pre-search - 1

§ What do the EPO Guidelines say about pre-search?

- A **pre-search algorithm** creates a list of documents to be inspected is created: automated search!

- Pre-search triggered by creation of European Search Report, European Search Opinion or Rule 62a and/or 63(1) EPC clarification request

Chapter IV – Search procedure and strategy

1. Procedure prior to searching

Upon creation of a European search report, a European search opinion or a clarification request under Rule 62a and/or 63(1), a ████████████ generating a list of documents to be inspected by the examiner is triggered. This creates a marker which serves as evidence in the file that the Search Division has started the search. The date of the start of the search is relevant for a possible refund of the search fee in case the application is withdrawn, refused or deemed to be withdrawn (see A-X, 10.2.1).

Background of pre-search - 2

§ **Primary objective**: to retrieve

- **Relevant prior art** under Article 54(2) EPC
- (Un)published co-pending applications,

  prior art under Article 54(3) EPC

§ **Secondary objective**: to gather information useful to the examiner

- CPC, IPC, FI/FT-classes potentially relevant for the search
- Potentially relevant terms/passages from the application
- *Work in progress*

§ Pre-search puts the examiner in a favourable position at the start of the search

§ Search can be re-focused after assessing pre-search results

- Pre-search increases the speed and quality of search

The pre-search algorithms

§ Several algorithms used!
  - By default, pre-search is fully automatic- no user input is needed
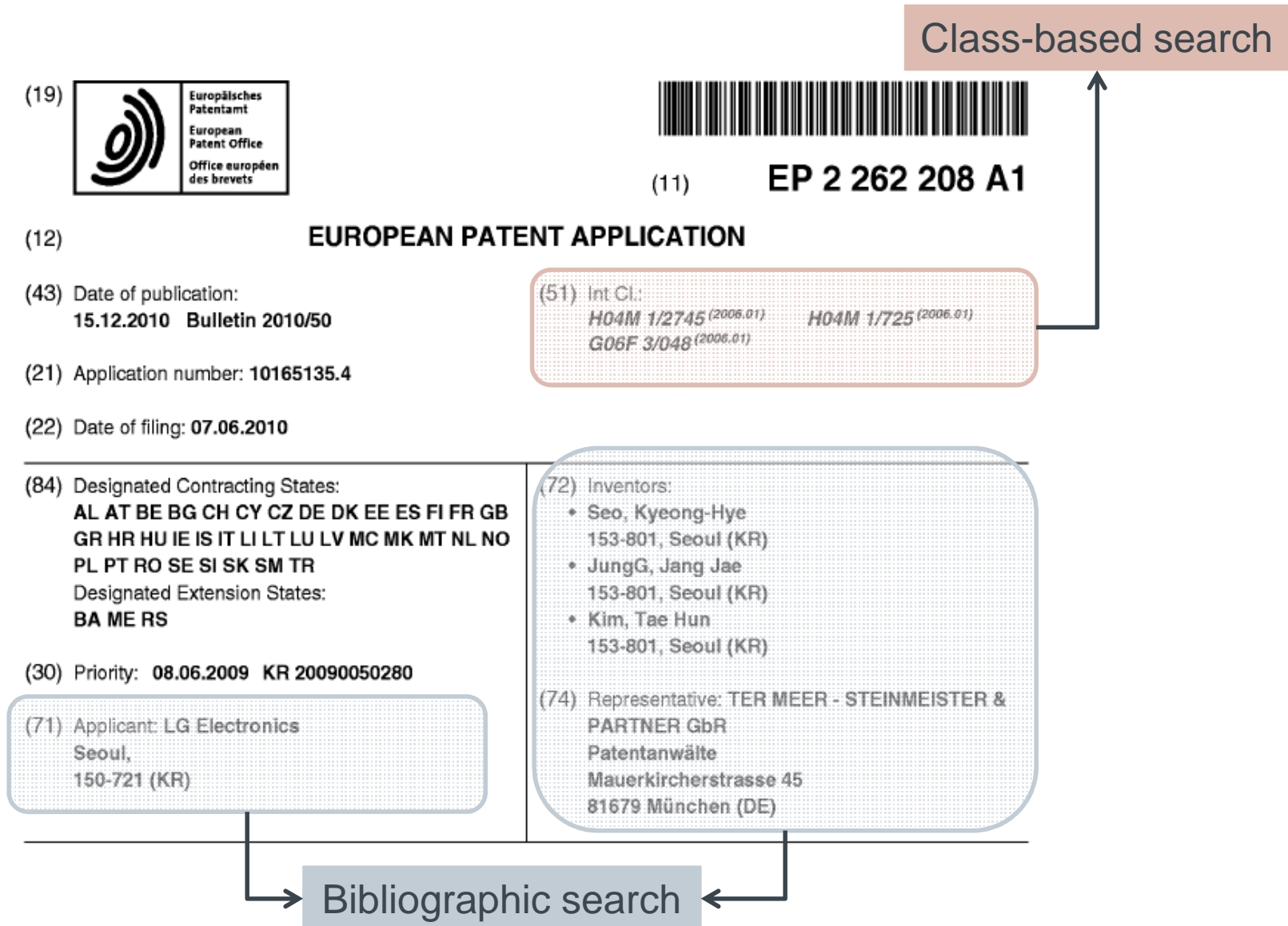§ Citation retrieval
  - Applicant citations
  - Citations from ISA
  - Citations from other Patent offices
  - Documents citing the application
  - Use of the One Portal Dossier for citations from the IP5 Offices
  - NPL citations included
§ Bibliographic search
§ Class-based search
§ Term-based search

Input for pre-search algorithms - 1

Class-based search

(11) **EP 2 262 208 A1**

EUROPEAN PATENT APPLICATION

(12)

(43) Date of publication:
15.12.2010   Bulletin 2010/50

(51) Int Cl.:
H04M 1/2745 (2006.01)   *H04M 1/725* (2006.01)
G06F 3/048 (2006.01)

(21) Application number: 10165135.4

(22) Date of filing: 07.06.2010

(84) Designated Contracting States:
AL AT BE BG CH CY CZ DE DK EE ES FI FR GB
GR HR HU IE IS IT LI LT LU LV MC MK MT NL NO
PL PT RO SE SI SK SM TR
Designated Extension States:
BA ME RS

(72) Inventors:
• Seo, Kyeong-Hye
  153-801, Seoul (KR)
• JungG, Jang Jae
  153-801, Seoul (KR)
• Kim, Tae Hun
  153-801, Seoul (KR)

(30) Priority: 08.06.2009 KR 20090050280

(71) Applicant: LG Electronics
Seoul,
150-721 (KR)

(74) Representative: TER MEER - STEINMEISTER &
PARTNER GbR
Patentanwälte
Mauerkircherstrasse 45
81679 München (DE)

Bibliographic search

Input for pre-search algorithms - 2

(54)   **Method for executing a menu in a mobile terminal and mobile terminal using the same**

(57)   The present disclosure is related to a method for executing a menu in a mobile terminal, the method comprising; inputting (S2) a drawing pattern (1-9) on a touch screen of the mobile terminal; displaying a menu corresponding to the drawing pattern and a sub menu thereof on the touch screen by comparing (S3) a pre-stored drawing pattern table with the drawing pattern; and executing (S6) the sub menu by selecting the sub menu and a mobile terminal thereof.

Term-based search

Citation retrieval

Patent documents cited in the description

*   KR 1020090050280 [0001]

FIG. 7
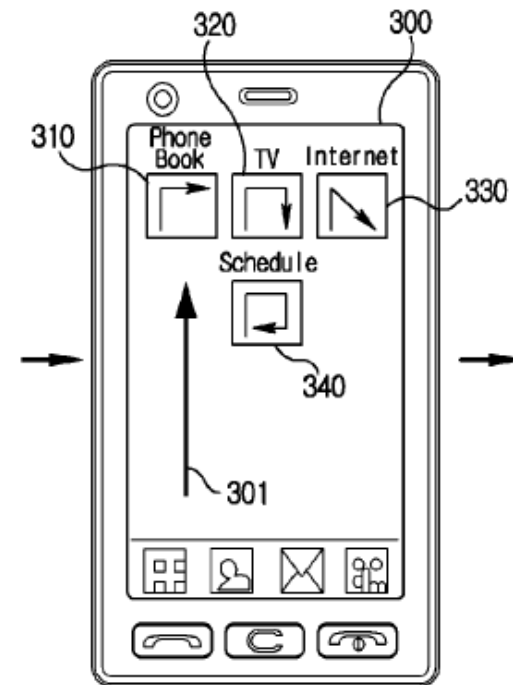
(b)

The bibliographic search

§ Implemented as the **APDEX** algorithm (developed by A. Materne)

§ Uses bibliographical information as search input
- Inventor names
- Applicant
- Representative

§ Will retrieve prior art from same applicant/inventors
- Co-pending (un)published applications
  - Article 54(3) EPC documents
- Documents relevant to the right to priority for the application

The class-based search

§ Implemented as the **FTRK** algorithm (developed by A. Materne)

§ Directed to Japanese prior art

§ Uses F-terms and FI-classes of any Japanese family member of the
application as search input

§ Will retrieve Japanese prior art having a Japanese classification similar to
the application

§ *Work in progress*:
  • Generalisation of class-based search
  • Extension to CPC-classes, IC-classes....

The term-based search

§ Implemented as the **Ansera-MLT, PS1** (developed by Y. Kingma) and **XFR** (developed by A. Materne) algorithms

§ Extracts terms or combination of terms from abstract, claims and/or description as search input

§ Will retrieve prior art disclosing these terms, *ranked* in an order of potential relevance

§ Ansera-MLT and PS1 extracts and searches for individual terms, and implements inverse document frequency ranking

§ XFR extracts and searches for combinations of terms, and implements Horváth-Materne ranking

Presentation of the results from pre-search - 1

§ Pre-search is triggered at the start of the search

§ When pre-search has finished (<5 minutes), the results (on average 47) are presented to the Examiner in the Viewer in a dedicated drawer

Citation from ISA

Document also found by Ansera-MLT

Citation from ISA

Presentation of the results from pre-search - 2

§ The pre-search results are ranked according to potential relevance:

Potential
Relevance

§ The examiner can therefore study the most relevant documents first

§ But how can we know which documents are the most relevant?

Evaluation of the pre-search algorithms - 1

§ After each run of pre-search, the publication numbers provided by pre-search are stored, as well as name of the algorithm(s) that found the document

§ When the examiner drafts the search report, the publication numbers of the documents cited are stored as well and **compared to the pre-search results**

§ The documents cited in the search report are the gold standard

§ This information enables us to evaluate
  • the efficiency of the pre-search algorithms
  • the ranking of the results of the pre-search algorithms

Evaluation of the pre-search algorithms - 2
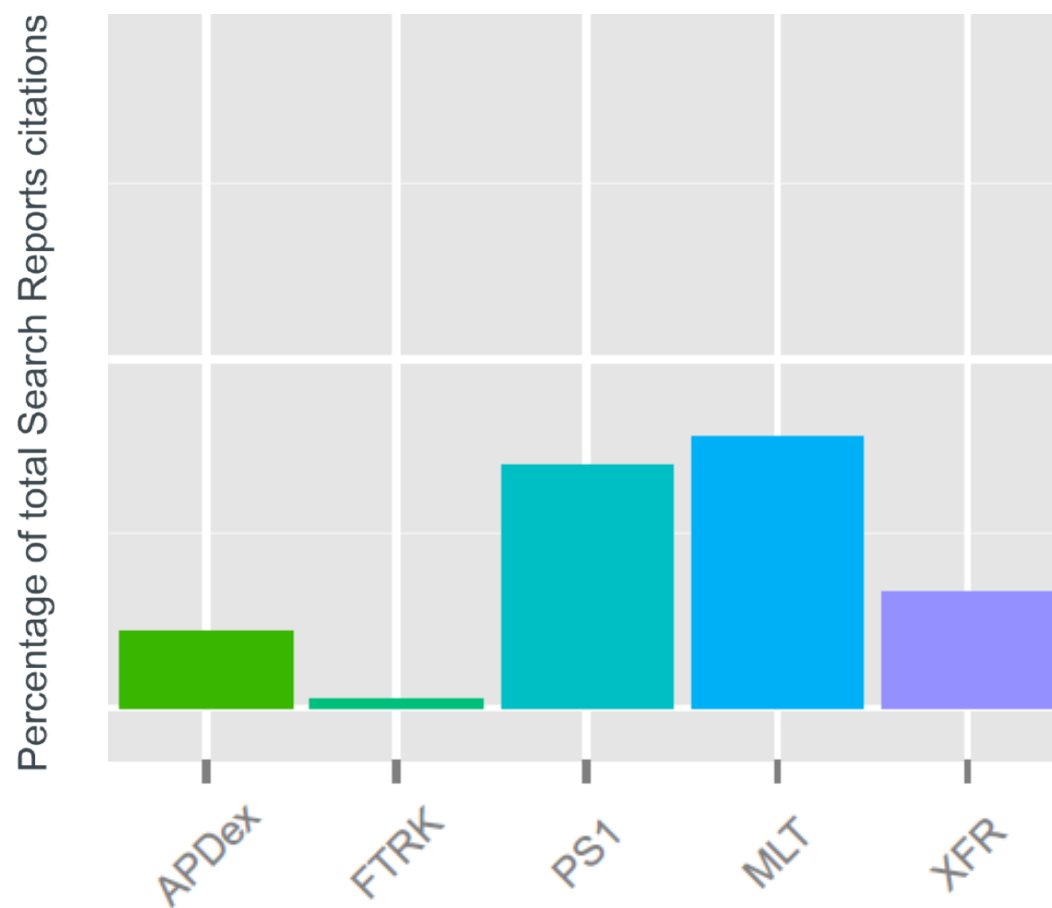
# Evaluation of the pre-search algorithms - 2
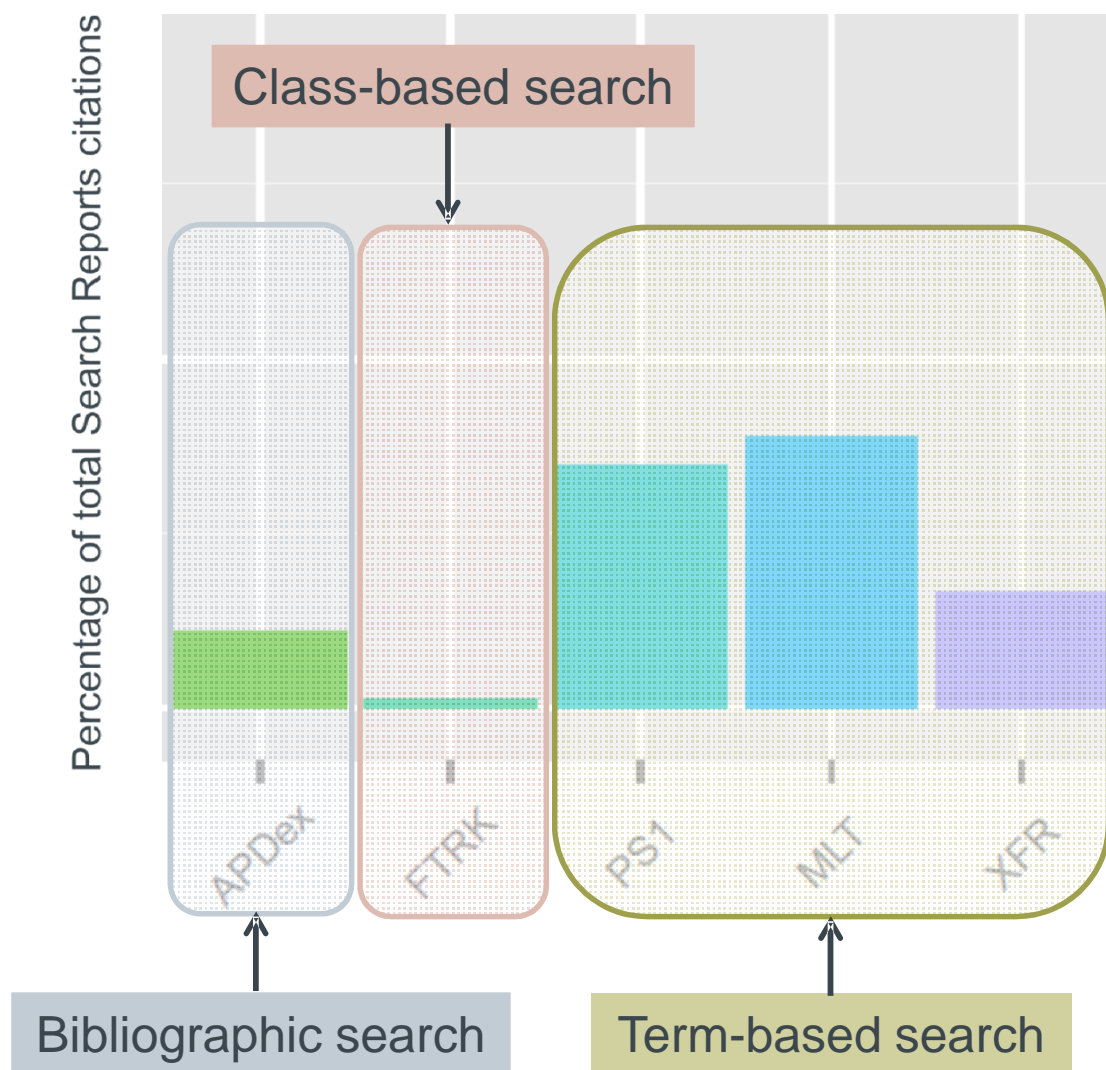
NB: Only patent publications taken into account for this evaluation!

Evaluation of the pre-search algorithms - 3

§ Evaluation of ranking for Ansera-MLT in pre-search- better ranked documents have higher chances of being cited in Search Report!

§ Proof of concept for pre-search ranking

Rank of document

(y-axis: Probability that document gets cited)
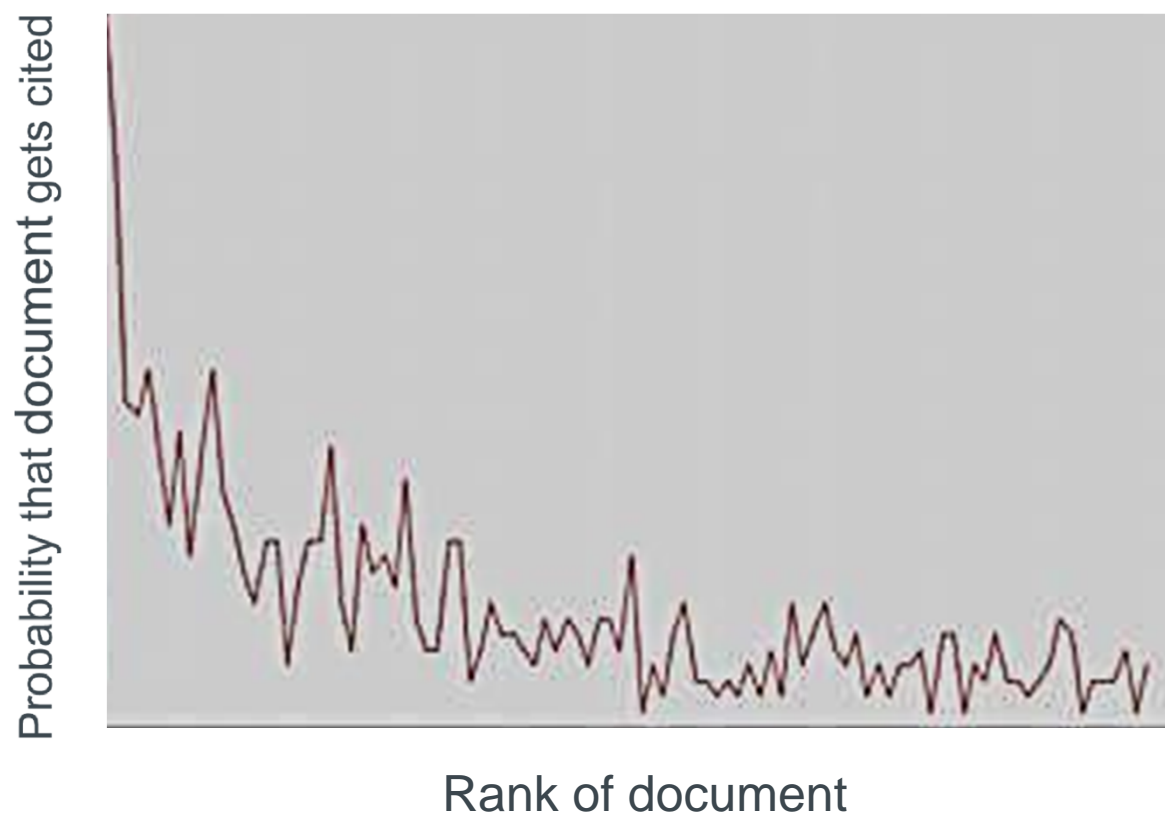
Working with results of an automated search - 1

§ Working with automated search tools might be challenging for the user

§ In the classical Boolean database search, the examiner "knows" why each document was included in the result set:
  • Example: Searching for all documents classified under G06F3/044/IC disclosing the term "proximity" in the WPI abstract

§ **Problem**: This does not apply for an automated search
  • "Why was this prima facie irrelevant document returned by pre-search?"
  • Might cause confusion and decreased trust in the automated search tools
  • Worst case scenario: User feels the need to study irrelevant document in more detail -> loss of time

Working with results of an automated search - 2

§ Proposed solutions based on **EPO experience with pre-search**:

- The user should have studied the application very carefully before evaluating any results from the automated search
  - Knowledge gives power to avoid wasting time on irrelevant prior art

- The users readily accept and appreciate automated search, but they will always be curious: "Why was this document returned?"

- The automated search tool should thus be able to inform the user in detail about the origin of each document in the result set**:**
  - Was the document cited by the applicant/ISA/USPTO?
    - Category of citation? For which claims?
  - Search terms used for term-bases search tools

Working with results of an automated search - 3

§ The usefulness of an automated search will depend on many factors, including

- Technical field of application
- Complexity of application
- The needs and taste of the individual user

§ Possible solution:

- Application-dependent settings for the automated search based on previous experience, "fine-tuning" - work in progress
- Give the users some control of pre-search settings

Working with results of an automated search - 4

§ How many documents should an automated search return?

- Depends on
  - Potential usefulness of the results
  - Time needed to study each document
    - Complexity of prior art
    - Functionality of document viewing software

§ Reasonable number: **50 documents** (**EPO experience with pre-search**)

- Balance between precision/recall of result
- The user should be able to control the amount of results

Working with results of an automated search - 5

§ Do the users accept/appreciate working with pre-search?
  - Some initial reluctance
  - Typical question a few years ago:
    - "Why do these documents appear in my working list?"

§ With more experience and knowledge, attitudes change
  - Typical question today:
    - "Why were no results returned from Ansera-MLT for my application?"

§ **Automated search via pre-search has become an integral part of the work of the EPO examiner**

Future of pre-search

§ Constant **improvements of algorithms**
  - The effect of any changes can be evaluated automatically

§ Optimize **number of documents returned** by pre-search
  - Quality of results estimated
    - Number of documents returned adjusted accordingly

§ Extending pre-search to **non-patent literature**
  - Highly important for certain technical fields

§ Improve **presentation** of results to user
  - Ensure that examiner is able to understand why documents were found by pre-search

§ Let pre-search provide more **additional information** to the user
  - CPC, IPC, FI/FT-classes potentially relevant for the search
  - Potentially relevant terms/passages from the application

Conclusions

§ The EPO pre-search framework provides the examiner with prior art found by several different state-of-the-art algorithms

§ The automatic evaluation of the efficiency of the algorithms puts the EPO in a favourable position to improve pre-search

§ Automatic search brings many benefits to the search professional
  • but some care should be taken when working with results from an automated search

§ Potential for future improvements is great
  • In the future, the role of the automated search will be even more important than today