Efficient Medical Information Retrieval in Encrypted Electronic Health Records

Cédric PRUSKI, François WISNIEWSKI
CRP Henri Tudor – CR SANTEC AHI, Luxembourg

Medical Informatics Europe MIE2012, Pisa, Italy, August 2012
Problem and motivation

- Implementation of an eHealth platform for the sharing of healthcare documents with a clear focus on security issues.
- To enhance the protection of patient privacy and prevent illegal usage, only de-identified and encrypted medical data are stored on the platform.
- Due to this configuration, documents content’s cannot be known without being fully decrypted.
- Impossible for the HPs to retrieve, in a efficient way, relevant information in this vast quantity of data for their daily practice.

Provide a solution to enhance information retrieval (IR) in encrypted EHR by preserving the patient privacy.
Agenda

1. Background

2. Method
   1. XDS Metadata and Meta-information extraction
   2. XDS Metadata and Meta-information exploitation

3. Example

4. Conclusions and perspectives
How to overcome the limitation caused by data encryption regarding IR?

- **Use of metadata to:**
  - Have a more precise indexation and representation of the content of documents stored in the platform
  - Provide a query enrichment mechanism in order to improve, in terms of relevance, the retrieval of medical documents

- **Constraints:**
  - The selection of the metadata shall not betray the identity of patients
  - The provided mechanisms shall be transparent for the end-users who are not familiar with IT
Method:
Metadata and Meta-information extraction

- Exploits the properties offered by standard metadata
  - XDS metadata: authorPerson, creationTime, classCode ...

- **Idea:** add an additional layer called Meta-information that describes the content of the document much precisely *without* revealing medical results and disclosing patient identity.

- **Proposal:** Provide a software application that extracts both the XDS metadata and the Meta-information by analysing the document before its encryption.
  - Use of NLP techniques and tools like MetaMap,
  - Use of vocabulary provided by standards termino-ontological resources like UMLS, SNOMED CT, LOINC, ICD-10 as Meta-information
### Example of metadata and meta-information

#### Meta-information Index

<table>
<thead>
<tr>
<th>Document data type for lab result in Hematology</th>
<th>UniqueDocID Reference</th>
<th>Encrypted symmetric key</th>
<th>XDS metadata</th>
</tr>
</thead>
<tbody>
<tr>
<td>Total Hemoglobin</td>
<td>12156121268431, 04153321652632, 12313115613935</td>
<td>Key 1</td>
<td>authorInstitution, authorPerson, authorRole, authorSpecialty comments, title, creationTime, formatCode eventCodeList + eventCodeDisplayNameList classCode + classCodeDisplayName typeCode + typeCodeDisplayName healthcareFacilityTypeCode + healthcareFacilityTypeCodeDisplayName practiceSettingCode + practiceSettingCodeDisplayName</td>
</tr>
<tr>
<td>Red Blood Cell Count RBC count</td>
<td>12156121268431, 12315646546512</td>
<td>Key 2</td>
<td>authorInstitution, authorPerson, authorRole, authorSpecialty comments, title, creationTime, formatCode eventCodeList + eventCodeDisplayNameList classCode + classCodeDisplayName typeCode + typeCodeDisplayName healthcareFacilityTypeCode + healthcareFacilityTypeCodeDisplayName practiceSettingCode + practiceSettingCodeDisplayName</td>
</tr>
<tr>
<td>Hematocrit Hct</td>
<td>04153321652632, 12315646546512</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Iron and Total Iron-binding Capacity</td>
<td>12313115613935</td>
<td>Key 2</td>
<td></td>
</tr>
<tr>
<td>Ferritin</td>
<td>12156121268431, 04153321652632</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>12313115613935</th>
<th>ABCDEF</th>
<th>Repository 2</th>
<th>SymKey 3</th>
</tr>
</thead>
<tbody>
<tr>
<td>...</td>
<td>...</td>
<td>...</td>
<td>...</td>
</tr>
</tbody>
</table>
Method:
Metadata and Meta-information exploitation

How to make the most out of the extracted metadata and meta-information to enhance IR in terms of relevance of the returned results?

- **Idea:** Enrich submitted queries with additional keywords taken from the set of XDS metadata and vocabulary provided by standard termino-ontological resources

- **Proposal:** Use an ontology that connect XDS metadata model with UMLS to enable extraction of additional keywords using ontological properties as well as keywords of the initial query.
  - Use of a lexical analyser to normalize terms of queries with the reference vocabulary (substitution of misspelled word synonyms, multilinguism, etc.)
  - The search space is reduced thanks to additional keywords
  - Gain in precision and recall
Meta-info index method
How to retrieve specific clinical information

1 - Queries specific Health data
2- Enriches the query
3- Submits the enriched query
4- Pre-selects a set of document with the XDS metadata
5- Refines the selection with the Meta information index
6- Retrieves the set of encrypted document
7- Decrypts the set of document and data mine the requested data.
8- Get the specific data
Conclusion

- Approach of Meta-information indexation that only tags the document, does not weaken security of the initial architecture (pseudonymisation and encryption of the document).

- Separation of Meta-information index from the XDS registry:
  - Prevent hacking of the Meta-information index to retrieve the patient pseudonyms
  - Speed up the processing of queries, gain in precision and recall

- Statistics to monitor the healthcare activity of the country could be done on the Meta-information.

- Work still needs to be validated with end users
- Work on maintenance of the ontology and the Meta-information index has to go on.
Thank you

François Wisniewski
francois.wisniewski@tudor.lu

Cédric Pruski
cedric.pruski@tudor.lu

CRP Henri Tudor – SANTEC AHI, Luxembourg