HEALTH BANK –
A Workbench for Data Science Applications in Healthcare

Hercules Dalianis
Clinical Text Mining Group
Department of Computer and Systems Sciences (DSV)
hercules@dsv.su.se
Clinical text mining group 2007-2014

Aron Henriksson, Mia Kvist, Sara Brissman, Maria Skeppstedt, Gunnar Nilsson, Martin Hassel, Sumithra Velupillai, Hideyuki Tanushi and Hercules Dalianis
(Not in photo Claudia Ehrentraut, and Rebecka Weegar)
Claudia Ehrentraut and Rebecka Weegar

HEALTH BANK
2 mil. patient records
2007-2014
Overview of talk

• Background
• Ethics
• The database
• The language in clinical text
• Applications for healthcare
• A workbench
• Collaborations
HEALTH BANK – The Swedish Health Record Research Bank
(Stockholm EPR Corpus)

- Karolinska University Hospital
  - TakeCare Intelligence
- First ethical permission 2008
- First database 2006-2008
- 4th database 2007-2014, ~ 2 million records
- And eight ethical permission!
Content in patient records

• Serial number, gender, age
• Admission, discharge date and time stamps
• Blood-, laboratory values, ICD-10 diagnosis codes
• Drugs - ATC-codes

**AND LOTS OF!!!**

Free text in Swedish
  - Physician’s notes, reasoning, nurses narratives, etc
Sensitive text

- In the free text under Social headings
  - Lots of personal names, phone numbers, addresses etc.
    - *The patient is assisted by her husband Anders, ph 070-567 32 55*
Medical language

- Noisy text
- Non-standard abbreviations
  - p5 - *pertrokatær femurfraktur* (hip fracture)
- Misspellings up to 10 percent of all words
  - *Parkisons, frctre*
- Missing subjects, incomplete sentences
- Physician expresses also herself very vague
  - *Possible infection, no signs of bacterias but*...
  - *No feber, no pain, no Parkisons*
Medicinskt journalspråk

Septisk pat, oklart fokus, rundodlas före Zinacef =>

Patienten har sepsis med oklart ursprung, bakterieodling tas från samtliga möjliga infektionsfokus, inklusive blododling, innan behandling med Zinacef inleds.
Medical language

Septicemic pat, unclear origin, roundcultured before Zinacef.

=>

The patient has septicemia of unclear origin, bacterial culture samples taken from all possible foci for infection, including blood culture samples, before commencing treatment with Zinacef.
ICD-10 coding

• 22 classes and 35,000 codes
  • (J18) Pneumonia, organism unspecified
    • (J18.0) Bronchopneumonia, unspecified
    • (J18.1) Lobar pneumonia, unspecified
    • (J18.2) Hypostatic pneumonia, unspecified
    • (J18.8) Other pneumonia, organism unspecified
    • (J18.9) Pneumonia, unspecified
Challenges in Healthcare – Adverse events

Healthcare-associated infections (HAI)

- International studies have found that up to 10 per cent of patients at any given time has healthcare associated infections, (Humphreys and Smyths, 2006)
- 10 percent or more of the in-patients in Europe obtain an HAI
- Three million injured patients and 50 000 deaths yearly only in Europe.
Adverse drug events (ADE)

- ADEs causes 3.7% of hospital admissions worldwide.
- One of the most common causes of death
- Seventh most common cause of death in Sweden
Diagnosis code assignment and validation

- 20 percent of the assigned ICD-10 diagnosis codes are erroneous
- Many codes to choose among
Early cancer symptoms

• The earlier cervical cancer, breast cancer or prostate cancer is detected the better

• Are there any early signs or early symptoms of cancer in the previous treatment of the patient?
Statistics and applications

- Costly to assign and validate ICD-10 diagnosis codes, over 20 percent of the assigned codes are erroneous
  => Automatic code assignment and validation using distributional semantics
- At least 10 percent of the admitted patients suffers from healthcare associated infections
  => Automatic detection and predictions of healthcare associated infections
- At least 4 percent of all admitted patient are admitted due to adverse drug events
  => Automatic detection of adverse drug events both known and unknown.
Statistics and applications (cont)

- Mining events preceding a cancer diagnosis.
  => Mining events in cervical cancer including negated events.
- Manual coding of pathology reports at Cancer Registry of Norway (Kreftregisteret) 25 full time coders and 180,000 reports yearly.
  => Automatic coding of pathology reports.
- Write discharge letter (epikris)
  => Automatic summarisation of health care episodes to a discharge letter.
Statistics and applications (cont)

• Patient/lay person reads his/her patient records
  => Automatic simplification of the patient record.

• Miss spellings and grammar errors in the patient record text.
  => Spelling and grammar checking of the patient record text.
Automatic code assignment and validation
(Unsupervised machine learning, random indexing - distributional semantics)

<table>
<thead>
<tr>
<th>Hosta (cough)</th>
</tr>
</thead>
<tbody>
<tr>
<td>J18.9 - Pneumoni, ospecificerad (Pneumonia, unspecified)</td>
</tr>
<tr>
<td>J15.9 - Bakteriell pneumoni, ospecificerad (Bacterial pneumonia, unspecified)</td>
</tr>
<tr>
<td>H66.9 - Mellanöreinflammation, ej specificerad som varig / icke varig</td>
</tr>
<tr>
<td>(Otitis media, unspecified)</td>
</tr>
<tr>
<td>J20.9 - Akut bronkit, ospecificerad, (Acute bronchitis, unspecified)</td>
</tr>
<tr>
<td>B34.9 - Virusinfektion, ospecificerad, (Viral infection, unspecified)</td>
</tr>
<tr>
<td>G96.9 - Sjukdom i centrala nervsystemet, ospecificerad</td>
</tr>
<tr>
<td>(Disorder of central nervous system, unspecified)</td>
</tr>
<tr>
<td>I50.9 - Hjärtinsufficiens, ospecificerad (Heart failure, unspecified)</td>
</tr>
<tr>
<td>F48.9 - Neurotiskt syndrom, ospecificerat (Neurotic disorder, unspecified)</td>
</tr>
<tr>
<td>C34.9 - Icke specificerad lokalisation av malign tumör i bronk &amp; lunga</td>
</tr>
<tr>
<td>(Bronchus or lung, unspecified)</td>
</tr>
<tr>
<td>L64.9 - Androgen alopeci, ospecificerad (Androgenic alopecia, unspecified)</td>
</tr>
</tbody>
</table>

**Fig. 10.7** Example in ICD-10 code suggestion (© 2009 The authors - reprinted with permission from the authors. Published in Dalianis et al. (2009).)
Manual annotation and training with machine learning (Supervised learning)

- Manual annotation or classification
  - symptoms, diagnosis, drugs and body parts
- Training with different machine learning algorithms
  - Recognizing patterns – create rules
Detect-HAI

- Detect-HAI analyses clinical text and data for Healthcare-associated infections
  - Based on 213 manually classified health care episodes (128 HAI/85 NON-HAI)
  - Both text 1,300,000 tokens and structured information
  - 93.7% recall and 79.7% precision using the Gradient Tree Boosting
Manually annotation for detection of clinical named entities

- Program modules for detection of
  - Symptom and diagnosis
  - Negation
  - Uncertainty
  - Period of time

76-årig kvinna med hypertoni och angina pectoris. Trolig hjärtinfarkt 2 år sedan. Inkommer med centrala bröstsmärtor utan utstrålning.

- 76-year old woman with hypertension and angina pectoris. Possible heart attack 2 years ago. Admitted to hospital with central chest pain without radiation.
Manually annotation for detection of clinical named entities

- Program modules for detection of
  - Symptom and diagnosis
  - Negation
  - Uncertainty
  - Period of time

76-årig kvinna med hypertoni och angina pectoris.
Trolig hjärtinfarkt 2 år sedan. Inkommer med centrala bröstsmärter utan utstrålning.

- 76-year old woman with hypertension and angina pectoris. Possible heart attack 2 years ago. Admitted to hospital with central chest pain without radiation.
Manually annotation for detection of clinical named entities

- Program modules for detection of clinical named entities
  - Symptom and diagnosis
  - Negation
  - Uncertainty
  - Period of time

76-årig kvinna med hypertoni och angina pectoris. Trolig hjärtinfarkt 2 år sedan. Inkommer med centrala bröstsmärtor utan utstrålning.

- 76-year old woman with hypertension and angina pectoris. Possible heart attack 2 years ago. Admitted to hospital with central chest pain without radiation.
Manually annotation for detection of clinical named entities

- Program modules for detection of
  - Symptom and diagnosis
  - Negation
  - Uncertainty
  - Period of time

76-årig kvinna med hypertoni och angina pectoris. Trolig hjärtinfarkt 2 år sedan. Inkommer med centrala bröstmärtor utan utstrålning.

- 76-year old woman with hypertension and angina pectoris. Possible heart attack 2 years ago. Admitted to hospital with central chest pain without radiation.
Manually annotation for detection of clinical named entities

- Program modules for detection of
  - Symptom and diagnosis
  - Negation
  - Uncertainty
  - Period of time

76-årig kvinna med hypertoni och angina pectoris. Trolig hjärtinfarkt 2 år sedan. Inkommer med centrala bröstsmärtor utan utstrålning.

- 76-year old woman with hypertension and angina pectoris. Possible heart attack 2 years ago. Admitted to hospital with central chest pain without radiation.
Manually annotation of internal medicine emergency unit records

- Over 8,000 annotated entities
  - 1,998 Disorders
  - 3,822 Findings
  - 1,539 Drugs
  - 750 Body parts
<table>
<thead>
<tr>
<th></th>
<th>Finding</th>
<th>Body_part</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>VRU: Vulva ua.</td>
<td>Vulva, no remarks, abbreviated</td>
</tr>
<tr>
<td>2</td>
<td>Vagina öppet 3 cm vid introitus, vaginaes längd 6 cm.</td>
<td></td>
</tr>
<tr>
<td>3</td>
<td>Finner en rel fast striktur som en ring belägen cirka 2 cm från portioplanet.</td>
<td></td>
</tr>
<tr>
<td>4</td>
<td>Vid manipulation lättblödande slemhinnor men inget tumörsuspekt.</td>
<td>easily-bleeding mucous membranes but not tumor suspicious</td>
</tr>
<tr>
<td>5</td>
<td>Kan inspekteras eller palperas.</td>
<td></td>
</tr>
<tr>
<td>6</td>
<td>Debridierar digitalt sammanväxtningar på den högra och vänstra sidan av vaginalväggen och kan därefter prova ut en vaginaldilatator 30 mm som genomsläpplig ut med hela vaginaes längd motsvarande 6 cm till uretramynningen från portioplanet.</td>
<td>vaginal wall</td>
</tr>
<tr>
<td></td>
<td></td>
<td>urethral orifice</td>
</tr>
<tr>
<td>7</td>
<td>Rekommenderas använda denna dilatator 3 gånger/ vecka tills att slemhinnorna slutar att blöda.</td>
<td>until mucous membranes stops bleeding</td>
</tr>
<tr>
<td>8</td>
<td>Får en återbesöktid till UT om cirka 4 månader.</td>
<td></td>
</tr>
<tr>
<td>9</td>
<td>Pat har fått en handskriven information med sig hem.</td>
<td></td>
</tr>
</tbody>
</table>
1. VRU: Vulva, no remarks, abbreviated

2. Vagina öppet 3 cm vid introitus, vagina längd 6 cm.

3. Finner en rel fast striktur som en ring belägen circa 2 cm från portioplanet.

4. Vid manipulation lättblödande slemhinnor men inget tumörsuspekt.

5. Kan inspekteras eller palperas.

6. Debridierar digitalt sammanväxtningar på den högra och vänstra sidan av vaginalväggen och kan därefter prova ut en vaginaldilatator 30 mm som genomsläpplig ut med hela vaginas längd motsvarande 6 cm till uretramynningen från portioplanet.

7. Rekommenderas använda denna dilatator 3 gånger/vecka tills att slemhinnorna slutar att blöda.

8. Får en återbesökstid till UT om circa 4 månader.

Table 7. The most frequent findings, disorders and negations found in the physicians’ notes.

<table>
<thead>
<tr>
<th>Most frequent findings and disorders</th>
<th>Nbr of instances</th>
<th>Most frequently negated findings and disorders</th>
<th>Nbr of negated instances</th>
<th>Findings and disorders with highest portion of negation</th>
<th>Portion negated</th>
</tr>
</thead>
<tbody>
<tr>
<td>cervixcancer (cervical cancer)</td>
<td>873</td>
<td>besvä (trouble/problem)</td>
<td>338</td>
<td>gynekologiska besvä (gynecological problems)</td>
<td>1.0</td>
</tr>
<tr>
<td>besvä (trouble/problem)</td>
<td>790</td>
<td>feber (fever)</td>
<td>243</td>
<td>palpabla resistenser (palpable resistance)</td>
<td>1.0</td>
</tr>
<tr>
<td>illamående (nausea)</td>
<td>677</td>
<td>illamående (nausea)</td>
<td>198</td>
<td>särskilda besvä (particular problems)</td>
<td>1.0</td>
</tr>
<tr>
<td>mår bra (feels well)</td>
<td>662</td>
<td>blödningar (bleedings)</td>
<td>171</td>
<td>nyttillkomna symtom (new symptoms)</td>
<td>0.96</td>
</tr>
<tr>
<td>smärta (pain)</td>
<td>656</td>
<td>smärta (pain)</td>
<td>150</td>
<td>nyttillkomna besvä (new problems)</td>
<td>0.92</td>
</tr>
<tr>
<td>tumör (tumor)</td>
<td>642</td>
<td>smärtor (pains)</td>
<td>126</td>
<td>infektionstecken (signs of infection)</td>
<td>0.89</td>
</tr>
<tr>
<td>smärtor (pains)</td>
<td>629</td>
<td>blödning (bleeding)</td>
<td>99</td>
<td>biljud (murmur)</td>
<td>0.86</td>
</tr>
<tr>
<td>feber (fever)</td>
<td>562</td>
<td>infektionstecken (signs of infection)</td>
<td>91</td>
<td>tumörstrukturer (tumor structures)</td>
<td>0.83</td>
</tr>
<tr>
<td>cancer (cancer)</td>
<td>508</td>
<td>tumör (tumor)</td>
<td>83</td>
<td>tumörsuspekta förändringar (tumor suspicious changes)</td>
<td>0.82</td>
</tr>
<tr>
<td>blödningar (bleedings)</td>
<td>491</td>
<td>nyttillkomna besvä (new troubles)</td>
<td>79</td>
<td>subjektiva besvä (subjective problems)</td>
<td>0.8</td>
</tr>
<tr>
<td>blödning (bleeding)</td>
<td>482</td>
<td>buksmärtor (pain of the abdomen)</td>
<td>72</td>
<td>tumörsuspekt (tumor suspicion)</td>
<td>0.80</td>
</tr>
<tr>
<td>skivepitelcancer (squamous cell carcinoma)</td>
<td>428</td>
<td>hydronefros (hydronephrosis)</td>
<td>65</td>
<td>spridning (spreading)</td>
<td>0.78</td>
</tr>
</tbody>
</table>
Kreftregisteret i Oslo
The Cancer Registry in Norway
Pathology report for breast cancer in Norwegian

Mammaresektat (ve. side) med infiltrerende ductalt karsinom, histologisk grad 3
Tumordiameter 15 mm
Lavgradig DCIS med utstrekning 4 mm I kranial retning fra tumor
Frie reseksjonsrender for infiltrerende tumor (3 mm kranialt)
Lavgradig DCIS under 2 mm fra kraniale reseksjonsrand
ER: ca 65 % av cellene positive
PGR: negativ
Ki-67: Hot-spot 23% positive celler. Cold spot 8%.
Gjennomsnitt 15%
HER-2: negativ
Tidl. BU 13:
3 sentinelle lymfeknuter uten påviste patologiske forandringer

⇒ Found diagnostic tests by a mockup system

Progesteronreceptorer: 1 (1 is a table value that corresponds to “negative”)
Østrogenrecepttorer: 4 (4 is a table value that corresponds to “65 %”)
KI67 Hotspot: 23 ∈ Samtidig Sentinell Node: 0
KI67 Gjennomsnitt hot/cold: 1 ∈ HER-2 Immunihistokjemi: 2 (table value)
Tumors histologiske grad : 3 ∈ Tumordiameter: 15

Courtesy by Rebecka Weegar
Translated to English:

Mamma specimen (ie. side) with infiltrating ductal carcinoma, histological grade 3
Tumor diameter 15 mm
Low-grade DCIS extending 4 mm in cranial direction from the tumor
Free resection margins for infiltrating tumor (3 mm cranially)
Low-grade DCIS less than 2 mm from the cranial resection margin

ER: ca 65 % of the cells are positive
PGR: negative
Ki-67: Hot-spot 23% positive cells. Cold spot 8%.
Average 15%
HER-2: negative
Prev. BU 13:

3 sentinel lymph nodes without proven pathological changes

Found concepts by a mockup system
Progesteronreceptor (PGR): 1 (1 is a table value that corresponds to “negative” in the text)
Samtidig Sentinell Node: 0
Østrogenrecepttor (ER): 4 (4 is a table value that corresponds to “65 %” in the text)
KI67 Hotspot: 23
Tumors histologiske grad (Histological grade): 3
KI67 Gjennomsnitt hot/cold (Average): 15
Tumordiameter (Tumor diameter): 15
(In some cases the data is not found in the text but in sketch attached to the pathology report).
Automatic summarisation of health care episodes to a discharge letter.

**Fig. 10.5** Example on automatic discharge summary creation. Redundant information is removed and high scoring information is added in the beginning of the summary from highest to lowest, low scoring information G, F and H, are excluded. Taken from Figure 3 in (Moen et al., 2016), (Licensed under Creative Commons.)
Data science applications in healthcare

- Automatic surveillance of healthcare-associated infections
- Detection and exploration of adverse drug events
- Diagnosis code assignment
- Text mining in the cancer domain
  - Cervical cancer - detect early symptoms
  - Pathology reports – extract diagnostic tests
- Text simplification of clinical narratives/discharge summary
- Text input, spell checking
- Comorbidity analysis
Comorbidity view

**Fig. 10.10** Screen shoot of Comorbidity-view which is a visualising tool for comorbidity network. It contains all the disorders in ICD-10 code form that patient records have been assigned. The data contains 605,587 patients from 2006-2008 from the Karolinska University hospital. The thicker the line in Comorbidity-view the more patients have both ICD-10 codes (Tanushi et al., 2011).
HEALTH BANK – the Swedish Health Record Research Bank

- An infrastructure
- A workbench
- A data exploration tool
- Give access to data to both researchers and industry
Researchers

- Epidemiologists
- Pharmacologists
- Medical researchers
  - To generate and evaluate hypotheses
- Data scientists
  - To build systems
Health Record Data Including sensitive text

Health Record Data Except sensitive text

HEALTH BANK

1) Viewing and analyzing aggregated data: Comorbidity view
   Drug view

2) Download un-aggregated health data including text

User without ethical permission

User with ethical permission
HEALTH BANK

- Contain all possible tools for visualisation of electronic patient records
- Contain all possible tools for processing clinical text
- For secondary use of data
Current academic users

- Sweden
  - Stockholm University
  - Karolinska Institutet
  - Karolinska University Hospital
  - Uppsala University
  - Gothenburg University
  - University of Borås
Abroad

- University of Turku, Finland
- University of Copenhagen and DTU- Danmarks Tekniske Universitet
- NTNU-Trondheim, Norway
- Vytautas Magnus University, Lithuania,
- UC San Diego and University of Utah, USA
- SAS Institute and Treat Systems, Denmark
New text book April 2018 open access

Hercules Dalianis

Clinical Text Mining

Secondary Use of Electronic Patient Records
Comments & Questions?
Hercules Dalianis

Clinical Text Mining

Secondary Use of Electronic Patient Records

Open access here:


Springer Open
### Medical Classifications and Terminologies

5.1 International Statistical Classification of Diseases and Related Health Problems (ICD) ........................................ 36
5.1.1 International Classification of Diseases for Oncology (ICD-O-3) ......................................................... 37
5.2 Systematized Nomenclature of Medicine: Clinical Terms (SNOMED CT) ..................................................... 37
5.3 Medical Subject Headings (MeSH) ........................................... 39
5.4 Unified Medical Language Systems (UMLS) ......................... 39
5.5 Anatomical Therapeutic Chemical Classification (ATC) ......... 40
5.6 Different Standards for Interoperability ................................. 41
5.6.1 Health Level 7 (HL7) .................................................... 41
5.6.2 OpenEHR .............................................................. 42
5.6.3 Mapping and Expanding Terminologies ......................... 42
5.7 Summary of Medical Classifications and Terminologies .......... 43

### Evaluation Metrics and Evaluation

6.1 Qualitative and Quantitative Evaluation ................................ 45
6.2 The Cranfield Paradigm .................................................... 45
6.3 Metrics ................................................................. 46
6.4 Annotation ............................................................. 48
6.5 Inter-Annotator Agreement (IAA) ...................................... 49
6.6 Confidence and Statistical Significance Testing .................... 50
6.7 Annotation Tools ...................................................... 51
6.8 Gold Standard ......................................................... 51
6.9 Summary of Evaluation Metrics and Annotation .................... 53
7 Basic Building Blocks for Clinical Text Processing ........................................... 55
  7.1 Definitions ......................................................................................... 55
  7.2 Segmentation and Tokenisation ....................................................... 56
  7.3 Morphological Processing ................................................................. 57
    7.3.1 Lemmatisation ............................................................................. 57
    7.3.2 Stemming ..................................................................................... 57
    7.3.3 Compound Splitting (Decompounding) ....................................... 58
    7.3.4 Abbreviation Detection and Expansion ..................................... 58
    7.3.5 Spell Checking and Spelling Error Correction ......................... 60
    7.3.6 Part-of-Speech Tagging (POS Tagging) ....................................... 63
  7.4 Syntactical Analysis ............................................................................. 64
    7.4.1 Shallow Parsing (Chunking) ...................................................... 65
    7.4.2 Grammar Tools ............................................................................ 65
  7.5 Semantic Analysis and Concept Extraction ....................................... 66
    7.5.1 Named Entity Recognition ......................................................... 67
    7.5.2 Negation Detection ..................................................................... 69
    7.5.3 Factuality Detection .................................................................... 71
    7.5.4 Relative Processing (Family History) ......................................... 72
    7.5.5 Temporal Processing .................................................................. 73
    7.5.6 Relation Extraction ...................................................................... 78
    7.5.7 Anaphora Resolution ................................................................... 80
  7.6 Summary of Basic Building Blocks for Clinical Text Processing ........ 81
## 8 Computational Methods for Text Analysis and Text Classification

- **8.1 Rule-Based Methods** ......................................................... 83
  - 8.1.1 Regular Expressions .................................................... 84
- **8.2 Machine Learning-Based Methods** .................................... 85
  - 8.2.1 Features and Feature Selection ..................................... 87
  - 8.2.2 Active Learning .......................................................... 89
  - 8.2.3 Pre-Annotation with Revision or Machine Assisted Annotation .......................................................... 90
  - 8.2.4 Clustering ................................................................ 91
  - 8.2.5 Topic Modelling ........................................................... 92
  - 8.2.6 Distributional Semantics .............................................. 92
  - 8.2.7 Association Rules ........................................................ 93
- **8.3 Explaining and Understanding the Results Produced** ............... 94
- **8.4 Computational Linguistic Modules for Clinical Text Processing** ................................................................................. 94
- **8.5 NLP Tools: UIMA, GATE, NLTK etc** ...................................... 95
- **8.6 Summary of Computational Methods for Text Analysis and Text Classification** ................................................................. 95

## 9 Ethics and Privacy of Patient Records for Clinical Text Mining Research

- **9.1 Ethical Permission** ............................................................. 97
- **9.2 Social Security Number** ...................................................... 98
- **9.3 Safe Storage** .................................................................. 99
- **9.4 Automatic De-Identification of Patient Records** ....................... 99
  - 9.4.1 Density of PHI in Electronic Patient Record Text ................. 101
  - 9.4.2 Pseudonymisation of Electronic Patient Records ............... 104
  - 9.4.3 Re-Identification and Privacy ......................................... 106
- **9.5 Summary of Ethics and Privacy of Patient Records for Clinical Text Mining Research** ......................................................... 107
<table>
<thead>
<tr>
<th>Section</th>
<th>Title</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>10.1</td>
<td>Detection and Prediction of Healthcare Associated Infections (HAIs)</td>
<td>109</td>
</tr>
<tr>
<td>10.1.1</td>
<td>Healthcare Associated Infections (HAIs)</td>
<td>109</td>
</tr>
<tr>
<td>10.1.2</td>
<td>Detecting and Predicting HAI</td>
<td>111</td>
</tr>
<tr>
<td>10.1.3</td>
<td>Commercial HAI Surveillance Systems and Systems in Practical Use</td>
<td>113</td>
</tr>
<tr>
<td>10.2</td>
<td>Detection of Adverse Drug Events (ADEs)</td>
<td>114</td>
</tr>
<tr>
<td>10.2.1</td>
<td>Adverse Drug Events (ADEs)</td>
<td>115</td>
</tr>
<tr>
<td>10.2.2</td>
<td>Resources for Adverse Drug Event Detection</td>
<td>116</td>
</tr>
<tr>
<td>10.2.3</td>
<td>Passive Surveillance of ADEs</td>
<td>118</td>
</tr>
<tr>
<td>10.2.4</td>
<td>Active Surveillance of ADEs</td>
<td>118</td>
</tr>
<tr>
<td>10.2.5</td>
<td>Approaches for ADE Detection</td>
<td>119</td>
</tr>
<tr>
<td>10.3</td>
<td>Suicide Prevention by Mining Electronic Patient Records</td>
<td>122</td>
</tr>
<tr>
<td>10.4</td>
<td>Mining Pathology Reports for Diagnostic Tests</td>
<td>124</td>
</tr>
<tr>
<td>10.4.1</td>
<td>The Case of the Cancer Registry of Norway</td>
<td>125</td>
</tr>
<tr>
<td>10.4.2</td>
<td>The Medical Text Extraction (Medex) System</td>
<td>126</td>
</tr>
<tr>
<td>10.5</td>
<td>Mining for Cancer Symptoms</td>
<td>128</td>
</tr>
<tr>
<td>10.6</td>
<td>Text Summarisation and Translation of Patient Record</td>
<td>129</td>
</tr>
<tr>
<td>10.6.1</td>
<td>Summarising the Patient Record</td>
<td>132</td>
</tr>
<tr>
<td>10.6.2</td>
<td>Other Approaches in Summarising the Patient Record</td>
<td>133</td>
</tr>
<tr>
<td>10.6.3</td>
<td>Summarising Medical Scientific Text</td>
<td>134</td>
</tr>
<tr>
<td>10.6.4</td>
<td>Simplification of the Patient Record for Laypeople</td>
<td>134</td>
</tr>
<tr>
<td>10.7</td>
<td>ICD-10 Diagnosis Code Assignment and Validation</td>
<td>136</td>
</tr>
<tr>
<td>10.7.1</td>
<td>Natural Language Generation from SNOMED CT</td>
<td>138</td>
</tr>
<tr>
<td>10.8</td>
<td>Search Cohort Selection and Similar Patient Cases</td>
<td>139</td>
</tr>
<tr>
<td>10.8.1</td>
<td>Comorbidities</td>
<td>139</td>
</tr>
<tr>
<td>10.8.2</td>
<td>Information Retrieval from Electronic Patient Records</td>
<td>141</td>
</tr>
<tr>
<td>10.8.3</td>
<td>Search Engine Solr</td>
<td>142</td>
</tr>
<tr>
<td>10.8.4</td>
<td>Supporting the Clinician in an Emergency Department with the Radiology Report</td>
<td>142</td>
</tr>
<tr>
<td>10.8.5</td>
<td>Incident Reporting</td>
<td>143</td>
</tr>
<tr>
<td>10.8.6</td>
<td>Hypothesis Generation</td>
<td>144</td>
</tr>
<tr>
<td>10.8.7</td>
<td>Practical Use of SNOMED CT</td>
<td>145</td>
</tr>
<tr>
<td>10.8.8</td>
<td>ICD-10 and SNOMED CT Code Mapping</td>
<td>146</td>
</tr>
<tr>
<td>10.8.9</td>
<td>Analysing the Patient’s Speech</td>
<td>147</td>
</tr>
<tr>
<td>10.8.10</td>
<td>MYCIN and Clinical Decision Support</td>
<td>147</td>
</tr>
<tr>
<td>10.8.11</td>
<td>IBM Watson Health</td>
<td>147</td>
</tr>
<tr>
<td>10.9</td>
<td>Summary of Applications of Clinical Text Mining</td>
<td>148</td>
</tr>
</tbody>
</table>
11 Networks and Shared Tasks in Clinical Text Mining
   11.1 Conferences, Workshops and Journals
   11.2 Summary of Networks and Shared Tasks in Clinical Text Mining

12 Conclusions and Outlook
   12.1 Outcomes

References

Index