MedNLP Task
at NTCIR 2012–13 & 2013–14

Mizuki MORITA, Yoshinobu KANO, Tomoko OHKUMA,
Mai MIYABE, Eiji ARAMAKI

2013-09-24 @ CLEFeHealth2013 (Valencia, Spain)
What is NTCIR MedNLP?

- Shared Task on NLP
- Medical Documents in Japanese
- 2012-
Who We Are?

- Organizers of NTCIR MedNLP Task

Mizuki MORITA  Yoshinobu KANO  Tomoko OHKUMA

Mai MIYABE  Eiji ARAMAKI
Outline

• NTCIR
• NLP in Medicine in Japan
• MedNLP in NTCIR
• NTCIR-10 MedNLP-1
• NTCIR-11 MedNLP-2
• Future: collaboration, tool sharing, etc.
• **NTCIR**
  • NLP in Medicine in Japan
  • MedNLP in NTCIR
  • NTCIR-10 MedNLP-1
  • NTCIR-11 MedNLP-2
  • Future: collaboration, tool sharing, etc.
IR Evaluation Forums

- TREC (1992-)
- **NTCIR (1997-)**
- CLEF (2000-)
- INEX (2002-)
- TRECVID (2003-)
- i2b2 (2004-)
- FIRE (2008-)
- ...

From Noriko KANDO
• Research infrastructure for evaluating information access

• Since 1997

• 1.5-year (18-month) cycle

• NTICR-11 in 2013/2014
Participants

<table>
<thead>
<tr>
<th>NTCIR</th>
<th># of active participant groups</th>
<th># of countries</th>
</tr>
</thead>
<tbody>
<tr>
<td>NTCIR-1</td>
<td>28</td>
<td>6</td>
</tr>
<tr>
<td>NTCIR-2</td>
<td>36</td>
<td>8</td>
</tr>
<tr>
<td>NTCIR-3</td>
<td>65</td>
<td>9</td>
</tr>
<tr>
<td>NTCIR-4</td>
<td>74</td>
<td>10</td>
</tr>
<tr>
<td>NTCIR-5</td>
<td>77</td>
<td>15</td>
</tr>
<tr>
<td>NTCIR-6</td>
<td>85</td>
<td>12</td>
</tr>
<tr>
<td>NTCIR-7</td>
<td>92</td>
<td>15</td>
</tr>
<tr>
<td>NTCIR-8</td>
<td>65</td>
<td>17</td>
</tr>
<tr>
<td>NTCIR-9</td>
<td>90</td>
<td>17</td>
</tr>
<tr>
<td>NTCIR-10</td>
<td>101</td>
<td>16</td>
</tr>
</tbody>
</table>

From Noriko KANDO
Tasks in NTCIR

From Noriko KANDO
Cycle at NTCIR round

1. Call for task proposals
2. Selection of task proposals by Committee
3. Call for task participant (registration)
4. Submission of task results
5. Task evaluation
6. NTCIR Conference
## NTCIR-11 Timeline

<table>
<thead>
<tr>
<th>Event</th>
<th>Date</th>
</tr>
</thead>
<tbody>
<tr>
<td>Call for task proposal</td>
<td>2013-03/05</td>
</tr>
<tr>
<td>Tasks were accepted</td>
<td>2013-06</td>
</tr>
<tr>
<td>Task participants registration due</td>
<td>2013-12</td>
</tr>
<tr>
<td>Formal run</td>
<td>2014-03/07</td>
</tr>
<tr>
<td>Results submission due</td>
<td>2014-08</td>
</tr>
<tr>
<td>NTCIR-11 Conference (Tokyo)</td>
<td>2014-12</td>
</tr>
</tbody>
</table>
Chairs in NTCIR-11

• **General co-chairs**
  – Noriko KANDO
  – Tsuneaki KATO
  – Douglas W. OARD
  – Tetsuya SAKAI
  – Mark SANDERSON

• **Program co-chairs**
  – Hideo JOHO
  – Kazuaki KISHIDA

**Founder (Mother?) of NTCIR**
Noriko KANDO
Tasks in NTCIR-11

• Core Tasks (6)
  – **IMine** - Search task and intent mining
  – **Math-2** - Mathematical information access
  – **MedNLP-2** - Medical NLP
  – **MobileClick** - Mobile information access
  – **RITE-VAL** - Recognizing inference in texts and validation
  – **SpokenQuery&Doc** - Spoken query and document

• Pilot Tasks (2)
  – **QALab** - QA lab for entrance exam
  – **Temporalia** - Temporal information access
For More Information

- Website: [http://research.nii.ac.jp/ntcir/](http://research.nii.ac.jp/ntcir/)

- Catch Noriko KANDO

Founder (Mother?) of NTCIR
Noriko KANDO
Outline

- NTCIR
- **NLP in Medicine in Japan**
  - MedNLP in NTCIR
  - NTCIR-10 MedNLP-1
  - NTCIR-11 MedNLP-2
- Future: collaboration, tool sharing, etc.
Current Medical NLP in Japan

• At very primitive level

http://blogs.yahoo.co.jp/super_honnzyurasuhakusyaku/62764775.html
Current Medical NLP in Japan

**Academia (NLP researchers)**
- has methods
- want to appeal it
- but no data

**Industry**
- has solution & budget
- want to make a social contribution
- but no data

**Hospital**
- has data
- but no solution & no budget
Barriers among them

1. Mobility of human resources is very low

2. Physicians don’t have experience outside of medicine
Barriers among them

1. Mobility of human resources is very low
   – Because of ‘**lifetime employment system**’ in Japan

2. Physicians don’t have experience outside of medicine
   – Because of **medical education system** in Japan
Medical Education Systems

- **Type 1** (Japan)
  - High School
  - Medical School

- **Type 2**
  - High School
  - Liberal Arts
  - Medical School

Physician
Need to Connect Them

Software

Industry

NTCIR

Hospital

Data

Method

Academia
Outline

- NTCIR
- NLP in Medicine in Japan
  - **MedNLP in NTCIR**
- NTCIR-10 MedNLP-1
- NTCIR-11 MedNLP-2
- Future: collaboration, tool sharing, etc.
Goal

• To create **community**
  – to meet, discuss, exchange ideas

• To support developing **practical tools**

• To increase the number of **researchers**
Medical Documents
“ACTOS was prescribed at St Luke’s Hospital, and his leg was swollen.”

De-identification
“ACTOS was prescribed at ■■■ Hospital, and his leg was swollen.”

NER (Named Entity Recognition)
“ACTOS was prescribed at ■■■ Hospital, and his leg was swollen.”

Normalization
“Pioglitazone was prescribed at ■■■ Hospital, and his leg was edema.”

Data Mining
“Ψ(Θ)(Θ)Ψ > & # 266 ; $ ※ ★?”
Medical Documents
“ACTOS was prescribed at St Luke’s Hospital, and his leg was swollen.”

De-identification
“ACTOS was prescribed at Hospital, and his leg was swollen.”

NER (Named Entity Recognition)
“ACTOS was prescribed at Hospital, and his leg was swollen.”

Normalization
“Pioglitazone was prescribed at Hospital, and his leg was edema.”

Data Mining
“Ψ(Θ)(Θ)Ψ > & # 8653; ; $ ※ $?”
Medical Documents
“ACTOS was prescribed at St Luke’s Hospital, and his leg was swollen.”

De-identification
“ACTOS was prescribed at Hospital, and his leg was swollen.”

NER (Named Entity Recognition)
“ACTOS was prescribed at Hospital, and his leg was swollen.”

Normalization
“Pioglitazone was prescribed at Hospital, and his leg was edema.”

Data Mining
“Ψ(Θ)(Θ)Ψ > & # ⬤ ; $ ⋆ ☛ ?”
Outline

- NTCIR
- NLP in Medicine in Japan
- MedNLP in NTCIR
- **NTCIR-10 MedNLP-1**
- NTCIR-11 MedNLP-2
- Future: collaboration, tool sharing, etc.
NTCIR-10 MedNLP-1

June 18-21, 2013 @ NII (Tokyo, Japan)
Strategy

Medical Documents
“ACTOS was prescribed at St Luke’s Hospital, and his leg was swollen.”

De-identification
“ACTOS was prescribed at [redacted] Hospital, and his leg was swollen.”

NER (Named Entity Recognition)
“ACTOS was prescribed at [redacted] Hospital, and his leg was swollen.”

Normalization
“Pioglitazone was prescribed at [redacted] Hospital, and his leg was edema.”

Data Mining
“ψ(θ)ψ > & # 2015 ; $ ※ ※ ?”
Subtasks

• Task 1: De-identification Task
  – To extract personal information

• Task 2: Complaint and Diagnosis Task
  – To extract patient’s medical information

• Task 3: Free Task
### Timeline for MedNLP-1

<table>
<thead>
<tr>
<th>Event</th>
<th>Date</th>
</tr>
</thead>
<tbody>
<tr>
<td>Registration open</td>
<td>2012-11-01</td>
</tr>
<tr>
<td>Registration close</td>
<td>2012-11-30</td>
</tr>
<tr>
<td>Share sample set</td>
<td>2012-12-10</td>
</tr>
<tr>
<td>Share test set</td>
<td>2013-01-25</td>
</tr>
<tr>
<td>Submission deadline</td>
<td>2013-02-01</td>
</tr>
<tr>
<td>NTCIR-10 conference</td>
<td>2013-06-18 / 21</td>
</tr>
</tbody>
</table>
Case history (in Japanese)

工場に勤めている64歳の男性。2025年月8月2日（来院5日前）頃から腹痛が生じるとともに、食欲不振、嘔気・嘔吐出現した。体幹は温かいが、末梢は湿潤冷汗でショック状態。明らかな運動麻痺はみられず。翌日、意識障害出現し、腎機能障害の増悪を認めて徐々に尿量低下し、8月9日18時10分に心肺停止。8月9日21時44分死亡確認。
工場に勤めている64歳の男性。2025年8月2日（来院5日前）頃から腹痛が生じるとともに、食欲不振、嘔気・嘔吐出現した。体幹は温かいが、末梢は湿潤冷汗でショック状態。明らかな運動麻痺はみられず。翌日、意識障害出現し、腎機能障害の増悪を認めて徐々に尿量低下し、8月9日18時10分に心肺停止。8月9日21時44分死亡確認。
Targeted Text

• Created (artificial) case history
  – Written by physicians
  – Totally 50 histories (3,365 sentences)

  – Sample set (for development)
    • 2,244 sentences

  – Test set (for evaluation)
    • 1,121 sentences
### Tags for Annotation

- **De-identification task (Task 1)**

<table>
<thead>
<tr>
<th>Type</th>
<th>Tag</th>
<th>Counts†</th>
</tr>
</thead>
<tbody>
<tr>
<td>Age</td>
<td>&lt;a&gt;</td>
<td>56</td>
</tr>
<tr>
<td>Time</td>
<td>&lt;t&gt;</td>
<td>355</td>
</tr>
<tr>
<td>Hospital name</td>
<td>&lt;h&gt;</td>
<td>75</td>
</tr>
<tr>
<td>Location</td>
<td>&lt;l&gt;</td>
<td>2</td>
</tr>
<tr>
<td>Person’s name</td>
<td>&lt;p&gt;</td>
<td>0</td>
</tr>
<tr>
<td>Sex</td>
<td>&lt;x&gt;</td>
<td>4</td>
</tr>
</tbody>
</table>

† Counts in the *sample set.*
Tags for Annotation

- Complaint and diagnosis task (Task 2)

<table>
<thead>
<tr>
<th>Type</th>
<th>Tag</th>
<th>Counts†</th>
</tr>
</thead>
<tbody>
<tr>
<td>Complaint and diagnosis</td>
<td>&lt;c&gt;</td>
<td>1,922</td>
</tr>
</tbody>
</table>

– Attributes

<table>
<thead>
<tr>
<th>Type</th>
<th>Attribute</th>
<th>Counts†</th>
</tr>
</thead>
<tbody>
<tr>
<td>Positive</td>
<td></td>
<td>1,314</td>
</tr>
<tr>
<td>Negation</td>
<td>modality=&quot;negation&quot;</td>
<td>504</td>
</tr>
<tr>
<td>Suspicion</td>
<td>modality=&quot;suspicion&quot;</td>
<td>72</td>
</tr>
<tr>
<td>Family</td>
<td>modality=&quot;family&quot;</td>
<td>32</td>
</tr>
</tbody>
</table>

† Counts in the sample set.
Participants

• Entry: 18 groups

<table>
<thead>
<tr>
<th>Domestic</th>
<th>Overseas</th>
</tr>
</thead>
<tbody>
<tr>
<td>15 groups</td>
<td>3 groups (USA, Taiwan)</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Academia</th>
<th>Industry</th>
</tr>
</thead>
<tbody>
<tr>
<td>12 groups</td>
<td>6 groups</td>
</tr>
</tbody>
</table>
Participants

- Submission: **12 groups**

<table>
<thead>
<tr>
<th>Domestic</th>
<th>Overseas</th>
</tr>
</thead>
<tbody>
<tr>
<td>10 groups</td>
<td>2 groups (USA, Taiwan)</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Academia</th>
<th>Industry</th>
</tr>
</thead>
<tbody>
<tr>
<td>7 groups</td>
<td>5 groups</td>
</tr>
</tbody>
</table>
Participants

- Submission: 12 groups†

<table>
<thead>
<tr>
<th>Task 1</th>
<th>Task 2</th>
<th>Task 3</th>
</tr>
</thead>
<tbody>
<tr>
<td>De-identification</td>
<td>Complaint &amp; diagnosis</td>
<td>Free</td>
</tr>
<tr>
<td>6 groups</td>
<td>11 groups</td>
<td>1 group</td>
</tr>
<tr>
<td>15 systems</td>
<td>25 systems</td>
<td>1 system</td>
</tr>
</tbody>
</table>

† Up to 3 systems per group.
# Methods Overview

<table>
<thead>
<tr>
<th>Group ID</th>
<th>Methods</th>
</tr>
</thead>
<tbody>
<tr>
<td>NTTD</td>
<td>Word match, Semi-supervised learning</td>
</tr>
<tr>
<td>KobeU</td>
<td>Structured perceptron</td>
</tr>
<tr>
<td>ulab</td>
<td>Online learning</td>
</tr>
<tr>
<td>msiknowledge</td>
<td>CRF*, Language Model</td>
</tr>
<tr>
<td>UT-FX</td>
<td>CRF</td>
</tr>
<tr>
<td>HCRL</td>
<td>CRF, Word match</td>
</tr>
<tr>
<td>niph</td>
<td>Word match</td>
</tr>
<tr>
<td>oka1</td>
<td>CRF, Word match</td>
</tr>
<tr>
<td>NECLA</td>
<td>CRF</td>
</tr>
<tr>
<td>cks01</td>
<td>CRF</td>
</tr>
<tr>
<td>SinicaNLP</td>
<td>Word match, Machine translation</td>
</tr>
</tbody>
</table>

* CRF: Conditional Random Field
## Methods Overview

<table>
<thead>
<tr>
<th>Group ID</th>
<th>Extra Resources</th>
</tr>
</thead>
<tbody>
<tr>
<td>NTTD</td>
<td>MEDIS Standard Masters (disease names)</td>
</tr>
<tr>
<td>KobeU</td>
<td></td>
</tr>
<tr>
<td>ulab</td>
<td>Japanese newspaper (disease names)</td>
</tr>
<tr>
<td>msiknowledge</td>
<td></td>
</tr>
<tr>
<td>UT-FX</td>
<td>MedDRA/J, MEDIS Standard Masters, Original corpus</td>
</tr>
<tr>
<td>HCRL</td>
<td>Japanese Wikipedia (disease names)</td>
</tr>
<tr>
<td>niph</td>
<td>Original dictionary</td>
</tr>
<tr>
<td>oka1</td>
<td></td>
</tr>
<tr>
<td>NECLA</td>
<td>UMLS, LSD</td>
</tr>
<tr>
<td>cks01</td>
<td>MEDIS Standard Masters (disease names)</td>
</tr>
<tr>
<td>SinicaNLP</td>
<td>Original dictionaries (in Chinese)</td>
</tr>
</tbody>
</table>
De-identification Task

Baseline system (CRF, MEDIS)
### De-identification Task

- **Performances for each tag**

<table>
<thead>
<tr>
<th>Type</th>
<th>Tag</th>
<th>Counts†</th>
<th>Best (F)</th>
<th>Average (F)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Age</td>
<td>&lt;a&gt;</td>
<td>56</td>
<td>93.6</td>
<td>85.6</td>
</tr>
<tr>
<td>Time</td>
<td>&lt;t&gt;</td>
<td>355</td>
<td>90.3</td>
<td>84.0</td>
</tr>
<tr>
<td>Hospital name</td>
<td>&lt;h&gt;</td>
<td>75</td>
<td>96.0</td>
<td>82.4</td>
</tr>
<tr>
<td>Location</td>
<td>&lt;l&gt;</td>
<td>2</td>
<td>100.0</td>
<td>6.66</td>
</tr>
<tr>
<td>Person’s name</td>
<td>&lt;p&gt;</td>
<td>0</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>Sex</td>
<td>&lt;x&gt;</td>
<td>4</td>
<td>100.0</td>
<td>80.0</td>
</tr>
</tbody>
</table>

† Counts in the sample set.
Complaint & Diagnosis Task

Baseline system (CRF, MEDIS)

2-way
Total

F-measure vs Systems
Complaint & Diagnosis Task

- Performances for each attribute

<table>
<thead>
<tr>
<th>Attribute</th>
<th>Counts†</th>
<th>Best (F)</th>
<th>Average (F)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>1,314</td>
<td>79.3</td>
<td>65.5</td>
</tr>
<tr>
<td>negation</td>
<td>504</td>
<td>77.8</td>
<td>64.8</td>
</tr>
<tr>
<td>suspicion</td>
<td>72</td>
<td>50.7</td>
<td>33.6</td>
</tr>
<tr>
<td>family</td>
<td>32</td>
<td>81.1</td>
<td>64.1</td>
</tr>
</tbody>
</table>

† Counts in the *sample set*. 
Comments

• Top 3 groups in Task 1 and 2 are identical.

• All these 3 groups are from industry.
Comments

• All top 3 groups used CRF and extra disease name dictionaries in Task 2.

• But the best group in Task 1 used word matching-based method.
Comments

• Surprises for us ...

• There were no Japanese members in one of the top 3 groups.

• Many groups (5/12) were from industry.
Outline

- NTCIR
- NLP in Medicine in Japan
- MedNLP in NTCIR
- NTCIR-10 MedNLP-1
- **NTCIR-11 MedNLP-2**
- Future: collaboration, tool sharing, etc.
Medical Documents
“ACTOS was prescribed at St Luke’s Hospital, and his leg was swollen.”

De-identification
“ACTOS was prescribed at ■■■ Hospital, and his leg was swollen.”

NER (Named Entity Recognition)
“ACTOS was prescribed at ■■■ Hospital, and his leg was swollen.”

Normalization
“Pioglitazone was prescribed at ■■■ Hospital, and his leg was edema.”

Data Mining
“Ψ(Θ)(Θ)Ψ > & # ☽ ; $ ※ ☾ ?”
Medical Documents
“ACTOS was prescribed at St Luke’s Hospital, and his leg was swollen.”

De-identification
“ACTOS was prescribed at Hospital, and his leg was swollen.”

NER (Named Entity Recognition)
“ACTOS was prescribed at Hospital, and his leg was swollen.”

Normalization
“Pioglitazone was prescribed at Hospital, and his leg was edema.”

Data Mining
“Ψ(Θ)(Θ)Ψ > & # ; $ ×∈?”
Task Settings

• **Subtask 1: Extracting Complaint and Diagnosis Task**
  • To extract patient’s medical information

• **Subtask 2: Normalizing Complaint and Diagnosis Task**
  • To normalize patient’s medical information
1994年8月11日頃より全身倦怠感、労作時の息切れを自覚。14日午前3時頃に黒色吐物を嘔吐したため救急車を要請し、当院救急外来を受診。診察中に吐血したため上部消化器出血を疑い、精査加療目的に緊急入院
Problems of Normalization (1/2)

マルケサーニ症候群
マルケサーニ症候群
マルケサーニ症候群
マルケサーニ症候群
マルケサーニ症候群
マルケサーニ症候群

ICD-10: Q871
Marchesani Syndrome
Problems of Normalization (2/2)

惡心増悪
胃のむかつき
恶心 (sickness)
食後恶心
嘔気 (vomiting)
吐気気
恶心阻 (nausea)

ICD-10 : R16
恶心及び嘔吐
(sickness & vomiting)
<table>
<thead>
<tr>
<th>Event</th>
<th>Date</th>
</tr>
</thead>
<tbody>
<tr>
<td>Registration open</td>
<td></td>
</tr>
<tr>
<td>Registration close</td>
<td>2013-12</td>
</tr>
<tr>
<td>Share test set</td>
<td>2014-01</td>
</tr>
<tr>
<td>Submission deadline</td>
<td>2014-08</td>
</tr>
<tr>
<td>NTCIR-10 conference</td>
<td>2014-12</td>
</tr>
</tbody>
</table>
Outline

• NTCIR
• NLP in Medicine in Japan
• MedNLP in NTCIR
• NTCIR-10 MedNLP-1
• NTCIR-11 MedNLP-2

• **Future: collaboration, tool sharing, etc.**
Kachako
Automation Platform+Toolkit
Kachako provides all of these

- end users: everything in GUIs
- power users: link with their systems
- sharing created components/workflows
- combination, execution and evaluation of tools
- creation and customization of tools

Kachako: fully-automated integrated NLP system

- GUIs/visualizers or APIs/command line calls
- Workflow generation
- Hadoop-UIMA bridge
- comparison, evaluation, statistics
  - feature tracer
  - workflow generator

- clustered Web service deployer
- NLP tools portable
- type hierarchy
- parallel process
  - Hadoop/HDFS
  - auto deployment, execution, monitor
- creation and customization of tools

- UIMA Framework
  - metadata interoperability
- Hadoop/HDFS
  - distributed processing

Yoshinobu Kano
Kachako Platform Overview

Component Repository
- Search UIMA components

Workflow
- Create UIMA workflow

Execution
- Deploy workflow

Analyze
- Analyze workflow results
Kachako Platform Overview

Component Repository
- Search UIMA components

Workflow
- Create UIMA workflow

Execution
- Deploy workflow

Analyze
- Analyze workflow results
Statistics of Comparisons

- many combinations x metrics x files
- sort by a metric, and get best combination
Generic Visualizer for Text

<table>
<thead>
<tr>
<th>Covered Text</th>
<th>begin</th>
<th>end</th>
<th>surfaceForm</th>
</tr>
</thead>
<tbody>
<tr>
<td>崩壊</td>
<td>722</td>
<td>724</td>
<td>崩壊</td>
</tr>
<tr>
<td>Cardiff</td>
<td>724</td>
<td>725</td>
<td>Cardiff</td>
</tr>
<tr>
<td>世界</td>
<td>726</td>
<td>728</td>
<td>世界</td>
</tr>
<tr>
<td>九州</td>
<td>728</td>
<td>729</td>
<td>九州</td>
</tr>
<tr>
<td>新</td>
<td>729</td>
<td>730</td>
<td>新</td>
</tr>
<tr>
<td>新</td>
<td>730</td>
<td>732</td>
<td>新</td>
</tr>
<tr>
<td>秩序</td>
<td>732</td>
<td>733</td>
<td>秩序</td>
</tr>
<tr>
<td>を求める</td>
<td>733</td>
<td>735</td>
<td>を求める</td>
</tr>
<tr>
<td>いる</td>
<td>735</td>
<td>736</td>
<td>いる</td>
</tr>
<tr>
<td>世界</td>
<td>736</td>
<td>740</td>
<td>世界</td>
</tr>
<tr>
<td>740</td>
<td>741</td>
<td>741</td>
<td>741</td>
</tr>
<tr>
<td>741</td>
<td>742</td>
<td>743</td>
<td>743</td>
</tr>
<tr>
<td>743</td>
<td>744</td>
<td>744</td>
<td>744</td>
</tr>
<tr>
<td>744</td>
<td>746</td>
<td>746</td>
<td>746</td>
</tr>
<tr>
<td>746</td>
<td>747</td>
<td>747</td>
<td>747</td>
</tr>
<tr>
<td>747</td>
<td>749</td>
<td>749</td>
<td>749</td>
</tr>
<tr>
<td>749</td>
<td>751</td>
<td>751</td>
<td>751</td>
</tr>
<tr>
<td>751</td>
<td>752</td>
<td>752</td>
<td>752</td>
</tr>
<tr>
<td>752</td>
<td>753</td>
<td>754</td>
<td>754</td>
</tr>
<tr>
<td>754</td>
<td>756</td>
<td>756</td>
<td>756</td>
</tr>
<tr>
<td>756</td>
<td>757</td>
<td>757</td>
<td>757</td>
</tr>
<tr>
<td>757</td>
<td>759</td>
<td>759</td>
<td>759</td>
</tr>
<tr>
<td>759</td>
<td>760</td>
<td>760</td>
<td>760</td>
</tr>
<tr>
<td>760</td>
<td>762</td>
<td>762</td>
<td>762</td>
</tr>
<tr>
<td>762</td>
<td>764</td>
<td>764</td>
<td>764</td>
</tr>
<tr>
<td>764</td>
<td>766</td>
<td>766</td>
<td>766</td>
</tr>
<tr>
<td>日本</td>
<td>766</td>
<td>767</td>
<td>日本</td>
</tr>
</tbody>
</table>
Kachako platform

- Kachako platform for full automation
  - workflow oriented system, compliant with UIMA
- Kachako provides automation features for
  - installation and execution
  - workflow creation
  - parallel distributed deployment
  - comparison and evaluation
  - machine learning (upcoming)
  - result analysis visualization
- Kachako ready-to-use compatible toolkit (next)
Toolkit and Collaborations

- Multi-lingual and Multi-modal components

  - Brain fMRI mining (U-Sussex/U Kochi Tech)
  - Image/Movie Recognition (U-Tokyo)
  - Voice Recognition / Synthesize (U-Nagoya)
  - Reading Eye Track (Rehab/Ninjal)
  - Basic NLP for EN, JP, ZH (own)

  - GeoNLP (NII/PRESTO)
  - Aphasia Support (National Rehab Center)

  - Machine Translation (NICT)

  - Toodai-Robot UT Exam (NII)
  - Patient Information Leaflet Mining (U-Kagawa/U-Kyushu)

  - BCCWJ (Ninjal)
  - Online Ad Optimization

Core Components

- Components by Collaborations

- Platform

- Yoshinobu Kano

- Question Answering (NTCIR/CMU/NII)
- MedNLP (U-Kyoto/NTCIR)
- BioNLP (own)
- Langrid (U-Kyoto)
Related Projects (by Kano)

- Medicine package leaflet (UK) comparison
  - PIL: Patient Information Leaflet
    - for patients
  - SPC: Summary of Product Characteristics
    - for doctors, experts

- Aphasia support system
  - help Aphasia people by personalized system
    - support language understanding and utterance
    - by NLP and image processing technologies
MedNLP Potential Futures

• Using Kachako toolkit,
  – MedNLP + Machine Translation
  – MedNLP + Question Answering
  – etc.

• What sort of collaboration possible/useful with you?
  – Cross-lingual? (Japanese resources<->European?)
Join NTCIR-11

– Meet again in Tokyo in Dec, 2014
Acknowledgements

• Participants

• NTCIR Project Office
  – Noriko KANDO
  – Tetsuya SAKAI
  – Kazuko MATSUO
  – Mariko OKADA
  – Others

• Physicians

• Annotators & Others
  – Hiroto IMACHI
  – Sachiko MASKAWA
  – Ayako YAMASHITA
  – Junko HIRANUMA