Fairview Research

A Year of Living Dangerously:
Commercial Annotation with Teamware in Asia

Matthew Petrillo

Fairview Research

FIG 2010

Sheffield, UK

The Dossier

Primary Mission:

Test remotely-managed manual annotation

Secondary Mission:

Develop commercial best practices

Duration:

1 Year

Personnel:

Lighthouse Intellectual Property Group

Team 1: Cheng Du, China

Team 2: Cebu, Philippines

Target:

Patent Documents

Equipment:

Teamware 1.x

"Known Unknowns"

Untested software

Untrained annotators

No formal practices

Language barriers

Complex text

Internet Bandwidth

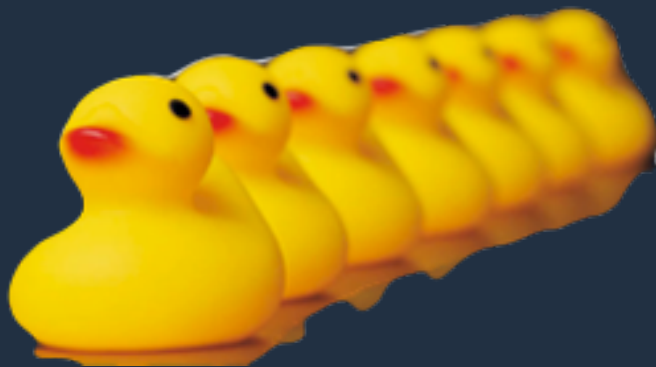Time & distance

Unforeseen Challenges

TECHNICAL:

Onsite technical support

Local hardware

Remote access issues

Teamware flexibility

Document presentation

PROCESS:

Post-annotation analysis

Generating "clean" corpora

Adapting academic practices

Guideline writing

MANAGEMENT:

Completion speed

Measuring and reporting

Managing down time / lag time

Trial Methodology

One month trial period

Single corpus broken into small sets for sequential tests

Documents double annotated, blind

Anonymous annotators, randomly selected by Teamware

Multiple testing rounds of increasing complexity

Each round = 1 week

Evaluation Criteria:

- Speed

- Agreement

- Consistency

Annotators unaware of criteria

```
<B740>
  <B741>
    <snm>Shanghai Jiaoda Patent Office -- WANG, Xi-lin; WANG, Gui-zhong</snm>
    <adr>
      <str/>
      <city/>
      <ctry>CN</ctry>
    </adr>
  </B741>
</B740>
</B700>
</sdobi>
<SDOAB>
  <abeng>
    <sec>
      <p>The invention claims tortula desertorum broth and method for rapidly increasing its microbiotic crusts in enviro
    </sec>
  </abeng>
</SDOAB>
```

Translated Chinese Patents mix the AGENT name with the address.

Problem: Western analysts could not distinguish place names correctly.

```
<B740>
  <B741>
    <snm>Shanghai Jiaoda Patent Office -- WANG, Xi-lin; WANG, Gui-zhong</snm>
    <adr>
      <str/>
      <city/>
      <ctry>CN</ctry>
    </adr>
  </B741>
</B740>
</B700>
</sdobi>
<SDOAB>
  <abeng>
    <sec>
      <p>The invention claims tortula desertorum broth and method for rapidly increasing its microbiotic crusts in enviro
    </sec>
  </abeng>
</SDOAB>
```

Agent	City	Province

Cheng Du Results Summary

| Trial	Dates	Corpus	Goal	# Docs Annotated / Avg. A/D*	Average Speed (per annotator, per document)	Est. weekly output (Team of 5)
Round 1	June 8 -12	2000 Chinese Patents (Bib & Abstract), in English	Differentiate between people, companies and addresses in 3 discrete segments	2000, double annotated 3 per doc.	2.5 minutes per annotator, per document	2100 – 2300 documents 6300 – 6900 annotations
Round 2	June 15 - 19	same as previous	Identify region in assignee address, add if non-existent	2000, double annotated 1 per doc.	1.8 minutes	4600 documents 9200 annotations
Round 3	June 21- July 3	500 Documents (US & EPO full text patents)	Edit, delete, or add measurement annotations	200, single annotated 30 per doc.	First set of 100: 46 minutes per annotator, per document	100-150 documents

*A/D = Annotations per document

Cheng Du, Round 1 Results

Accuracy & Consistency
IAA F1 Measures over 4 test sets

IAA Results held steady for Round 2
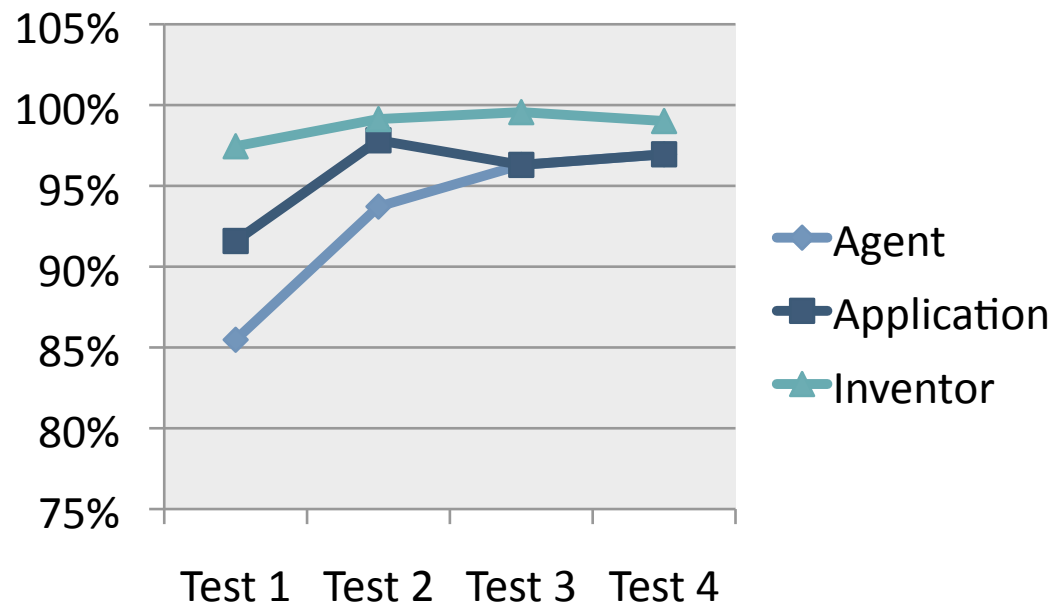
Speed:
Set 1: 20 minutes per doc.
Set 2: 6.9 minutes per doc.
Set 4: 2.5 minutes per doc.

Round 2: 1.8 minute per doc.

Fairview Research

Measuring curation activity on pre-annotated full-text patents

Cheng Du, Round 3: Results
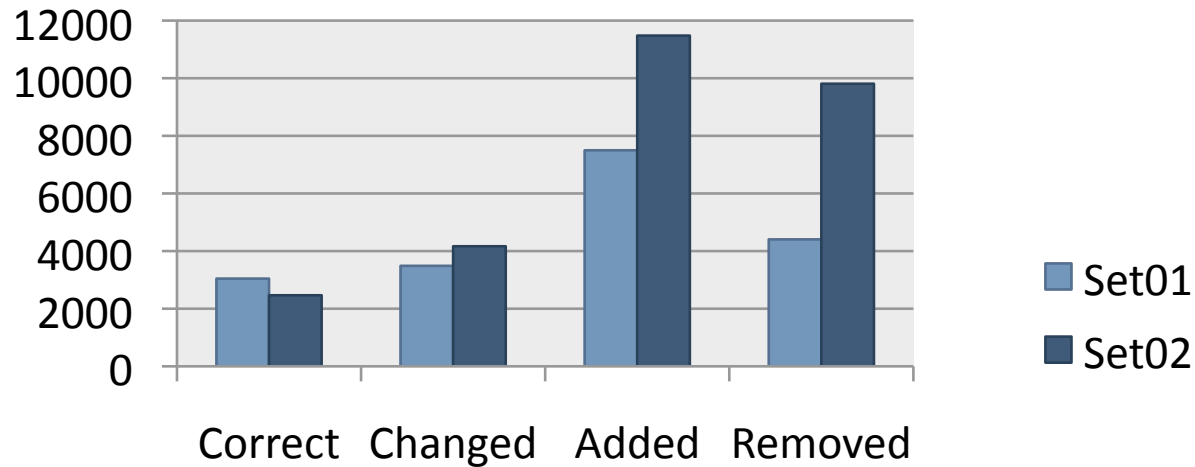
Speed:

46 minutes per doc.

Output:

Avg. ~700 actions per annotator / day
Avg. 100-200 Documents per week

Paperwork is Like Death and Taxes

Lighthouse Annotations Trials

Detailed Project Plan

Project Set-Up Phase May 17 – 29

Objectives:	Status:	Notes:	
Identify target documents Responsible: MP, CD	Complete	Agreed on 1000 CN translated patents provided by Lighthouse	
Identify annotation goals Responsible: MP	Complete	Identified structural inconsistencies between CN patents and Alexandria data normalization. Two areas to target: Inventor, Applicant, and Agent	
Create detailed project overview and Annotator instructions Responsible: MP, AR	Complete	Detailed instructions delivered	
Create training Slides showing steps for annotators Responsible: MP, JS	Complete	Delivered to Christa	

Installation Phase – June 1-5

Objectives:	Status:	Notes:	
Set up servers in China based on specifications provided by Sheffield. Responsible: LH China	Complete as of June 4	Initial installation in China had wrong version of Java. Issue resolved.	
Install Teamware on China Servers Responsible: MA, KK, RS	Complete as of June 5	Installation completed by Milan Agonotovic with assistance from Robbert Smet in NL and Kamassan Kaisiepo in China	

- Annotator Guidelines

- Training Materials

- Detailed Project Planners

- Project Objectives

- Generic Annotator Guidelines

- Annotator Evaluations

- Spreadsheets for stats

- Naming conventions for projects, corpora

Lessons Learned from Cheng Du

PROCESS

- Significant set-up and post-annotation evaluation time
- Training requirements surprisingly low for simple tasks
- Start prepping new projects as soon as annotators start
- Better evaluation tools and corpus management tools needed

MANAGEMENT

- Active management of projects is necessary and time-consuming
- Annotators can be very fast and can remain consistent with clearly-defined repetitive tasks
- F1 Measure is not adequate for evaluating curation tasks
- Aggregate IAA data does not highlight trouble spots
- Randomization does not control for potential problems between pairs

TECHNICAL

- Close to 100 Teamware-specific issues were identified and addressed in one-month test period

Cebu Trials: All Curation

Fairview Research

"ExoPatent" Documents

Three Rounds:
Round 1: Measurement, without features
Round 2: Measurement, with features
Round 3: Dosage form, with features

Simplified trial format

Cebu Results Summary

| Trial	Dates	Corpus	Goal	# Docs Annotated / Avg. A/D*	Average Speed (per annotator, per document)	Est. weekly output (Team of 5)
Round 1	July 20 -24 In three sets.	101 ExoPatent Documents	Edit, remove and add measurement annotations.	101	1 hr, 3 minutes in Set 1. 30.8 minutes by Set 3.	About 175
Round 2	August 3-14 Round 2 repeated, Aug. 10-14	100 ExoPatent Documents	Edit, remove and add measurement annotations.	100	20 minutes, 43 seconds	About 840
Round 3	Aug 17 - 21	100 ExoPatent Documents	Edit, remove and add Dosage Form annotations.	100	20 minutes, 28 seconds	About 875

*A/D = Annotations per document

Cebu Results: Annotator Speed

Fairview Research
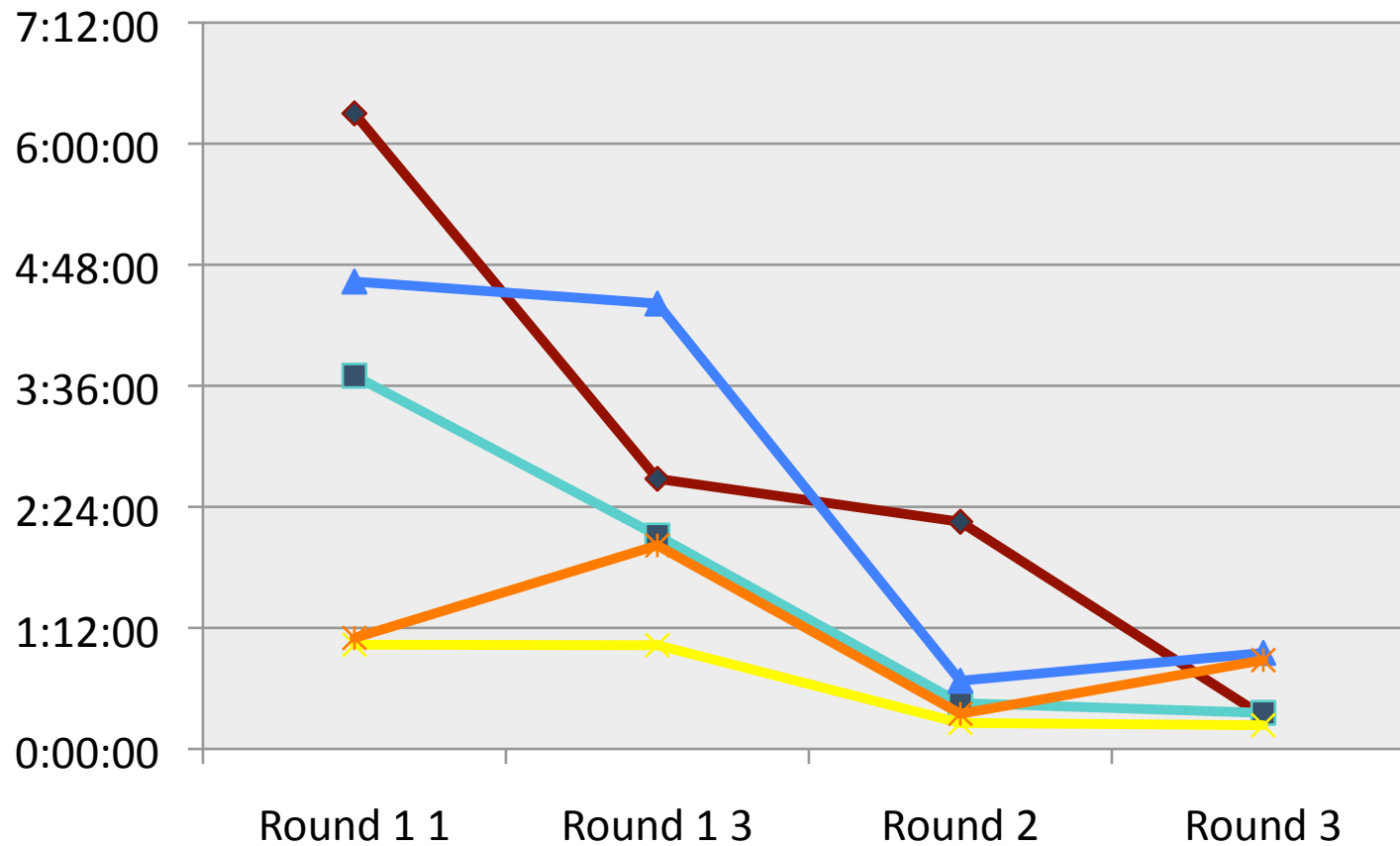
Typical Time per Annotator, per Document
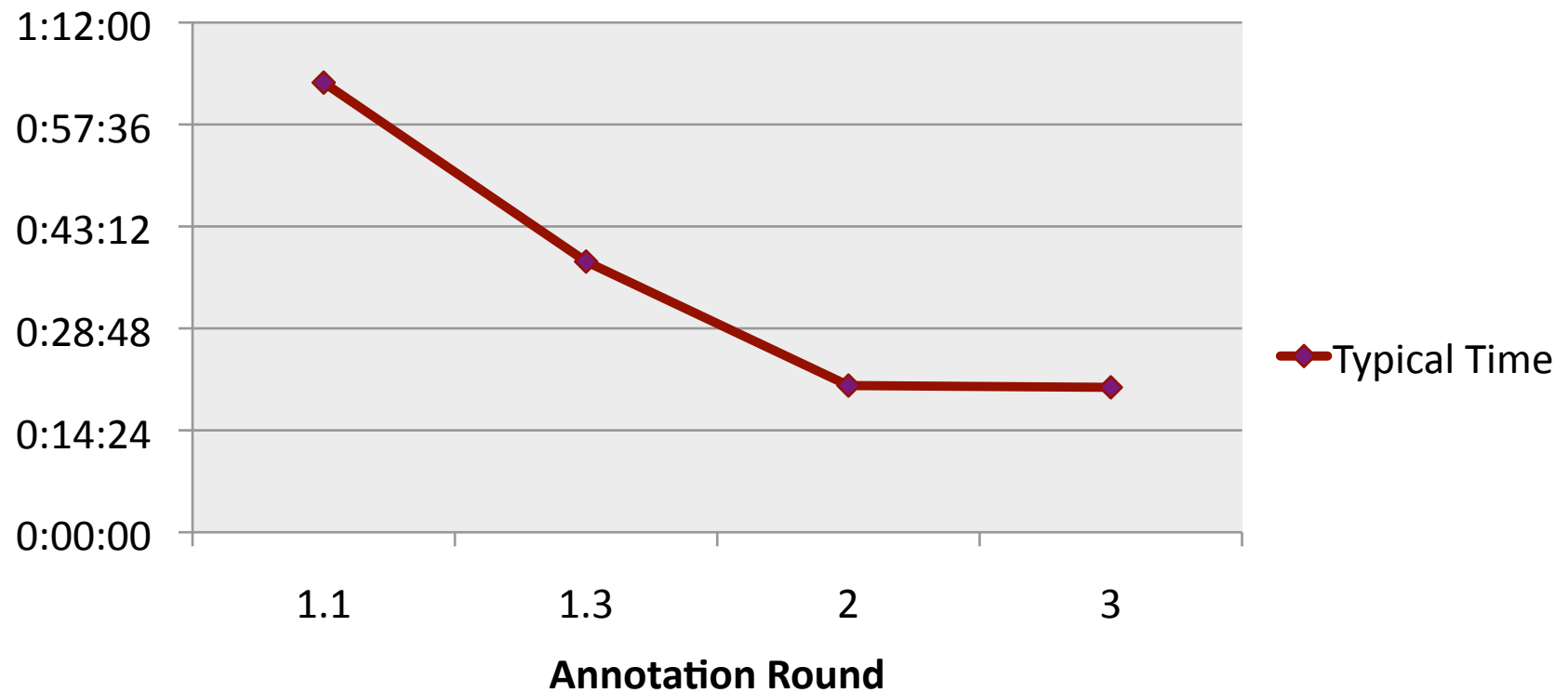
Comparing Annotators

Total Time Per Annotator, per Round

Managing "Correctness"

machine (purple) is correct; annotator 1 is inconsistent; annotator 2 is consistent but incorrect

Example from Cebu Round 1 Annotator Evaluation

Curating the Curators

"two or more" should not be annotated. remove if machine annotates it; find the inconsistency?

methdilazine hydrochloride, mebhydroline napadisylate, mequitazine, cyproheptadine hydrochloride, clemastine fumarate, epinastine hydrochloride, olopatadine hydrochloride, fexofenadine hydrochloride, loratadine, bepotastine besilate, mizolastine, and NIP-531. These can be used either individually or in combination of two or more. [0018] The amount of the antihistamine is determined based on the type and amount of other medicines used in combination according to a known formulation technique. A dose for an adult is usually 1-300 mg/day, and preferably 3-150 mg/day.

[0019] Examples of the analeptic include caffeine and sodium benzoate, caffeine, anhydrous caffeine, dl-methylephedrine hydrochloride, dl-methylephedrine saccharin salt, ephedrine hydrochloride, phenylpropanolamine hydrochloride, phenylephrine, l-methylephedrine hydrochloride, methoxyphenamine hydrochloride, dl-epinephrine hydrochloride, dl-isoproterenol hydrochloride, isoproterenol sulfate, orciprenaline sulfate, terbutaline sulfate, salbutamol sulfate, trimethoquinol hydrochloride, hexoprenaline sulfate, clorprenaline hydrochloride, tulobuterol hydrochloride, procaterol hydrochloride, pirbuterol hydrochloride, fenoterol hydrobromide, formoterol fumarate, clenbuterol hydrochloride, mabuterol hydrochloride, ethylcysteine hydrochloride, methylcysteine hydrochloride, and pseudoephedrine. These can be used either individually or in combination of two or more. [0020] The amount of the analeptic is determined based on the type and amount of other medicines used in combination according to a known formulation technique. A dose for an adult is usually 1-900 mg/day, and preferably 5-300 mg/day.

[0021] Examples of the galenical include powders of galenicals such as ephedra herb, nandina fruit, Japanese cherry bark, polygala root, glycyrrhiza, platycodon root, apricot kernel, plantago seed, plantago herb, lycoris, senega, kudzuvine root, fritillary bulb, gambir, fennel, scutellaria root, trichosanthes seed, cinnamon bark, oriental bezoar, schisandra fruit, asiasarum root, tatarian aster root, musk, codonopsis root, ginger, mulberry bark, perilla herb, panax rhizome, citrus unshiu peel, ginseng, ophiopogon tuber, and pinellia tuber, and their extracts. These can be used either individually or in combination of two or more. [0022] The amount of the galenical is determined based on the type and amount of other medicines used in combination according to a known formulation technique. A dose for an adult is usually 0.001-300 g/day (as an extract, converted into the raw galenical) or 0.0004-60 g/day (as a powder), and preferably 0.005-10 g/day (as an extract, converted into the raw galenical) or 0.001-3 g/day (as a powder). [0023] Examples of the antacid or mucosal protective agent include aminoacetic acid, magnesium oxide, magnesium carbonate, magnesium silicate, synthetic aluminum silicate, synthetic hydrotalcite, dihydroaluminum aminoacetate, aluminum hydroxide gel, aluminum hydroxide dried gel, aluminum hydroxide-magnesium carbonate mixture co-dried gel, aluminum hydroxide-sodium hydrogencarbonate coprecipitate, aluminum hydroxide-calcium carbonate-magnesium carbonate coprecipitate, magnesium hydroxide-aluminum potassium sulfate coprecipitate, magnesium aluminometasilicate, magnesium hydroxide, aluminum sulfate, and sucralfate. These can be used either individually or in combination of two or more. [0024] The amount of the antacid or mucosal protective agent is determined based on the type and amount of other medicines used in combination according to a known formulation technique. A dose for an adult is usually 10-8,000 mg/day, and preferably 100-4,000 mg/day.

Cebu Trial: Lessons Learned

Fairview Research

- Speed improvements were consistent with Cheng Du results

- "Accuracy" is best measured as a function of consistency

- Carefully develop guidelines to ensure comparable results

- Annotators will be confronted for different rules for different types of annotations, even when applying a "simplest possible" annotation rule.

- On highly subjective tasks, plan for extra time to accommodate several rounds of review to overcome ambiguities and improve consistency.

- Subsequent annotation sets on the same document should go faster. Thus, one annotation type per document may require 20 minutes, but two may require only 25 minutes.

- The Philippine annotators performed better when able to ask questions directly and when the goal of the annotation (not just the goal of the trial) was explained thoroughly.

- Incentive systems need to be evaluated to produce desired results

Overall Results

Remotely-managed annotation projects feasible

Standardized project documentation

More accurate project time and cost estimates

Provable quality measures

Annotator incentives to support project goals

Upgrades to GATE evaluation tools

Major improvements in GATE Teamware

Two teams "certified" to perform annotation work

Recent Contributions

Gold Standard for Measurement annotations

Extended Measurement annotation pipeline

Improved product demonstrators

Pharmaceutical annotation set developed in conjunction with Ontotext (GPCRs, Parameters, molecular measures, etc.)

Alignment experiments

Experimentation with Medline and UMLS annotations:
- •Anatomical structures
- •Diseases

Participation in EU Projects with GATE Team

To Do

Continue to refine and learn

Better automatic systems for annotator evaluation

Teamware 2.0

Annotation with "experts" group

"Round the World" annotation project

Integration with end-user search systems

THANK YOU!

Questions?