

Appraise |ə'prāz|

Acceptable Can easily be fixed None of both Flag Error

Christian Federmann, DFKI LT Lab

Appraise |ə'prāz|

verb [with obj.]

assess the value or quality of

✓ Acceptable

✎ Can easily be fixed

✗ None of both

! Flag Error

Christian Federmann, DFKI LT Lab

Overview

- ▶ Appraise in a nutshell
- ▶ Annotation tasks
- ▶ Added extras
- ▶ Available for free
- ▶ Outlook

```
# -*- coding: utf-8 -*-  
"""  
Project: Appraise evaluation system  
Author: Christian Federmann <cfedermann@dfki.de>  
"""  
import logging  
from django.contrib.auth.views import login as LOGIN, logout as LOGOUT  
from django.shortcuts import render  
from appraise.settings import LOG_LEVEL, LOG_HANDLER, COMMIT_TAG  
  
# Setup logging support.  
logging.basicConfig(level=LOG_LEVEL)  
LOGGER = logging.getLogger('appraise.views')  
LOGGER.addHandler(LOG_HANDLER)  
  
def frontpage(request):  
    """  
    Renders the front page view.  
    """  
    LOGGER.info('Rendering frontpage view for user "{0}".'.format(  
        request.user.username or "Anonymous"))  
  
    dictionary = {  
        'commit_tag': COMMIT_TAG,  
        'title': 'Appraise evaluation system',  
    }  
  
    return render(request, 'frontpage.html', dictionary)  
  
def login(request, template_name):  
    """  
    Renders login view by connecting to django.contrib.auth.views.
```

Appraise in a nutshell

System Description

- ▶ **Appraise** is...
 - ▶ an open-source tool for MT evaluation
 - ▶ available under a BSD license from GitHub
 - ▶ actively maintained and extended
 - ▶ alive and kicking since 2008...

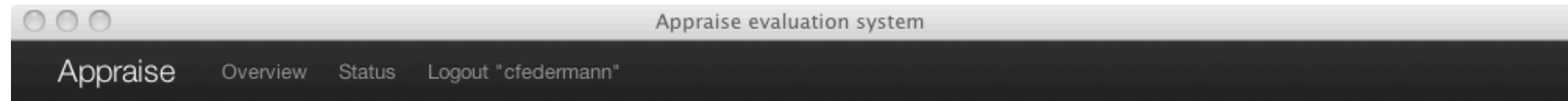
Features

- ▶ **Appraise** implements...
 - ▶ translation quality checking
 - ▶ ranking of translations
 - ▶ error classification
 - ▶ manual post-editing

Batteries Included

- ▶ **Appraise** comes...
 - ▶ ready for deployment as web-based service
 - ▶ with inter-annotator agreement scores
 - ▶ offering XML-based import/export formats
 - ▶ featuring full-blown administration backend

Say “hello”...



Welcome to Appraise!

This is the Appraise evaluation system for manual evaluation of machine translation output. It supports collaborative collection of human feedback for evaluation tasks such as *Translation Quality Checking*, *Ranking and Error Classification*, and *Manual Post-Editing*.

[See evaluation tasks »](#)

This is the GitHub version [7fe29f1a](#) of the Appraise evaluation system. Some rights reserved.

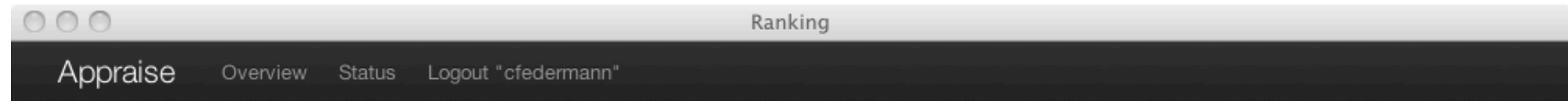
This is the GitHub version [7fe29f1a](#) of the Appraise evaluation system. Some rights reserved.

Annotation Tasks

Ranking

- ▶ **Ranking** of $n \geq 2$ translations
 - ▶ source text and, optionally, reference shown
 - ▶ wherever available, context is displayed
 - ▶ standard browser controls for ranking
 - ▶ UI supports keyboard shortcuts

Ranking



007/403

Schwergängiger Motor **Kundendienst anfordern** E 23
— Source

Rank 1 Rank 2

Customer Service Request
— Translation 1

Rank 1 Rank 2

After-sales service request
— Translation 2

This is the GitHub version [7fe29f1a](#) of the Appraise evaluation system. Some rights reserved.

3-Way Ranking

Appraise Overview Status Logout "cfedermann"

001/200

Thanks to grant funding (subsidy of 2.8 million) they were able to purchase equipment and for the period of two years provide qualified personnel and operation. The European Union announces having allocated in 2010 2.2 thousand million Euros, a good token of observance of its promise of 7.2 thousand million Euros in 2012.
— Source

Dank dieser Finanzierung (Zuschüsse in Höhe von 2,8 Mio. CZK) konnten die notwendigen Einrichtungen und die Ausstattung beschafft werden und für die Dauer von 2 Jahren ist für qualifiziertes Personal und für den Betrieb gesorgt. Die EU verkündet, im Jahre 2010 2,2 Milliarden Euro freigesetzt zu haben, was sie auf den richtigen Weg zur Respektbezahlung ihres Engagements von 7,2 Milliarden im Jahre 2012 bringt.
— Reference

Dank Subventionsfinanzierung (Subvention von 2.8 Millionen) konnten sie Ausstattung kaufen und sorgen für den Zeitraum von zwei Jahren für qualifiziertes Personal und Operation.
— Translation A

Dank Subventionsfinanzierung (Subvention von 2.8 Millionen) waren sie fähig, um Ausstattung zu kaufen, und sorgen für den Zeitraum von zwei Jahren für qualifiziertes Personal und Operation.
— Translation B

A > B A = B A < B

Error Classification

- ▶ **Classification** of errors in given translation
 - ▶ source (or target) text and translation shown
 - ▶ classification based on (Vilar et al., 2006)
 - ▶ global errors (“missing words”, “too many”)
 - ▶ local error classes (“minor”, “severe”)

Error Classification

Appraise Overview Status Logout "cfedermann"

001/1364

WILO PowerPoint-Präsentation WILO Präsentationsvorlage Version 1.0;<NewLine/>
— Source

WILO PowerPoint-Präsentation
— Translation

Missing words Too many errors

WILO	Click to classify errors for this word...	
PowerPoint-Präsentation	Terminology:	<input checked="" type="radio"/> None <input type="radio"/> Minor <input type="radio"/> Severe
	Lexical choice:	<input checked="" type="radio"/> None <input type="radio"/> Minor <input type="radio"/> Severe
	Syntax (ordering):	<input checked="" type="radio"/> None <input type="radio"/> Minor <input type="radio"/> Severe
	Insertion (extra word):	<input checked="" type="radio"/> None <input type="radio"/> Minor <input type="radio"/> Severe
	Morphology:	<input checked="" type="radio"/> None <input type="radio"/> Minor <input type="radio"/> Severe
	Misspelling:	<input checked="" type="radio"/> None <input type="radio"/> Minor <input type="radio"/> Severe
	Punctuation:	<input checked="" type="radio"/> None <input type="radio"/> Minor <input type="radio"/> Severe
	Other (idiom, etc.):	<input checked="" type="radio"/> None <input type="radio"/> Minor <input type="radio"/> Severe
	Click to classify errors for this word...	

Error Summary

No errors classified yet...

This is the GitHub version [7fe29f1a](#) of the Appraise evaluation system. Some rights reserved.

This is the GitHub version [11659179](#) of the Appraise evaluation system. Some rights reserved.

Quality Estimation

- ▶ **Quickly estimate** quality of given translations
 - ▶ source text and translation shown
 - ▶ depending on task design, reference given
 - ▶ user selects one of three result categories
 - ▶ focus on quick decisions (“gut feeling”)

Quality Estimation

The screenshot shows a web browser window titled "Translation Quality Checking". The interface includes a navigation bar with "Appraise", "Overview", "Status", and "Logout 'cfedermann'". A light blue header bar displays "002/063". Below this, two entries are shown: "Bilanz **Ausstehende Einlagen , davon eingefordert** sonstige immaterielle Vermögensgegenstände" (Source) and "Unbezahlte fordert gezeichneten Kapitals" (Translation). At the bottom, there are four buttons: "Acceptable", "Can easily be fixed", "None of both", and "Flag Error". A footer note states: "This is the GitHub version [7fe29f1a](#) of the Appraise evaluation system. Some rights reserved."

Translation Quality Checking

Appraise Overview Status Logout "cfedermann"

002/063

Bilanz **Ausstehende Einlagen , davon eingefordert** sonstige immaterielle Vermögensgegenstände
— Source

Unbezahlte fordert gezeichneten Kapitals
— Translation

✓ Acceptable ✎ Can easily be fixed ✘ None of both ⚠ Flag Error

This is the GitHub version [7fe29f1a](#) of the Appraise evaluation system. Some rights reserved.

Post-editing

- ▶ **Post-edit** a given translation
 - ▶ source and one or more translations shown
 - ▶ user selects translation “easiest to post-edit”
 - ▶ “Translate from scratch” mode
 - ▶ simple standard browser controls used

Post-editing

Post-editing

Appraise Overview Status Logout "cfedermann"

002/008

Die Sprache der Originalbetriebsanleitung ist Deutsch. **Alle weiteren Sprachen dieser Anleitung sind eine Übersetzung der Originalbetriebsanleitung.** Die Einbau- und Betriebsanleitung ist Bestandteil des Produktes.
— Source

Edit translation 1
All other languages of these instructions are translations of the original operating instructions.
— Translation 1

Edit translation 2
All further languages of these instructions are a translation of the original operating instruction.
— Translation 2

Edit translation 3
All further languages of this guidance are a translation of the original manual.
— Translation 3

Translate from scratch

Added Extras

Task Status Overview

»Holy Diver« Status Overview

User name	Overall completion	Average duration
cfedermann	<div style="width: 100%;"><div style="width: 100%;"></div></div> 5/5	4.03 sec
johndoe	0/5	0.00 sec
janedoe	0/5	0.00 sec
Combined	0/5	4.03 sec

Results Overview

A<B	A=B	A>B
2 (40.00%)	2 (40.00%)	1 (20.00%)

Administration Tools

[📄 Export results](#) [👤 Agreement data](#)

Agreement Scores

- ▶ **Appraise** gives you...
 - ▶ Krippendorff's α (*Krippendorff, 2004*)
 - ▶ Fleiss' κ (*Fleiss, 1971*) based on (*Cohen, 1960*)
 - ▶ Bennett's S (*Bennett, 1954*)
 - ▶ Scott's π (*Scott, 1955*)

Agreement Scores

»TermEval-COLING« Status Overview

User name	Overall completion	Average duration
Tobias.Wunner	<div style="width: 100%; background-color: green;">142/142</div>	12.13 sec
Noemi.Aeppli	<div style="width: 100%; background-color: green;">142/142</div>	59.05 sec
Thierry.Declerck	<div style="width: 100%; background-color: green;">142/142</div>	21.55 sec
Mihael.Arcan	<div style="width: 100%; background-color: green;">142/142</div>	20.86 sec
Paul.Buitelaar	<div style="width: 100%; background-color: green;">142/142</div>	25.45 sec
cfedermann	<div style="width: 100%; background-color: green;">142/142</div>	6.59 sec
Combined	<div style="width: 100%; background-color: green;">142/142</div>	24.27 sec

Results Overview

ACCEPTABLE	CAN_EASILY_BE_FIXED	NONE_OF_BOTH
504 (59.15%)	250 (29.34%)	98 (11.50%)

Inter-annotator Agreement

Krippendorff's Alpha	Fleiss' Kappa	Bennett's S	Scott's Pi
0.35544494026774454	0.3566810487909947	0.46690140845070416	0.354687531266884

0.35544494026774454	0.3566810487909947	0.46690140845070416	0.354687531266884
Krippendorff's Alpha	Fleiss' Kappa	Bennett's S	Scott's Pi

Inter-annotator Agreement

0.35544494026774454	0.3566810487909947	0.46690140845070416	0.354687531266884
---------------------	--------------------	---------------------	-------------------

Admin Backend

Change EvaluationTask object | Django site admin

Django administration Welcome, cfedermann. Change password / Log out

Home > Evaluation > EvaluationTask objects > <evaluation-task id="74">

Change EvaluationTask object History View on site

Required Information

Task name:
Unique, descriptive name for this evaluation task.

Task type: **Ranking**
Type choice for this evaluation task.

Task XML source: **source-xml/aaa.xml**
XML source file for this evaluation task.

Optional Information

Active?
Indicates that this evaluation task is still in use.

Random order?
Indicates that items from this evaluation task should be shown in random order.

Description:
(Optional) Text describing this evaluation task.

Users:

Available users	Chosen users
<input type="text"/> euroscript-CZEN	Select your choice(s) and click <input type="button" value="➔"/> taraxue-admin

Available for free

Hosted on GitHub



Hosted on GitHub



► **Get it at...**

► <https://github.com/cfedermann/Appraise>

Outlook

Outlook

- ▶ **Appraise** will...
 - ▶ happily support people trying it out
 - ▶ soon get more refined task status views
 - ▶ explore new annotation user interfaces
 - ▶ improve from your feedback/pull requests

Questions?

Acknowledgements

This work has been funded under the Seventh Framework Programme for Research and Technological Development of the European Commission (grant agreement no.: 249119) through the T4ME contract.

We are grateful to the anonymous reviewers for their valuable feedback on the draft version of the MTM paper.