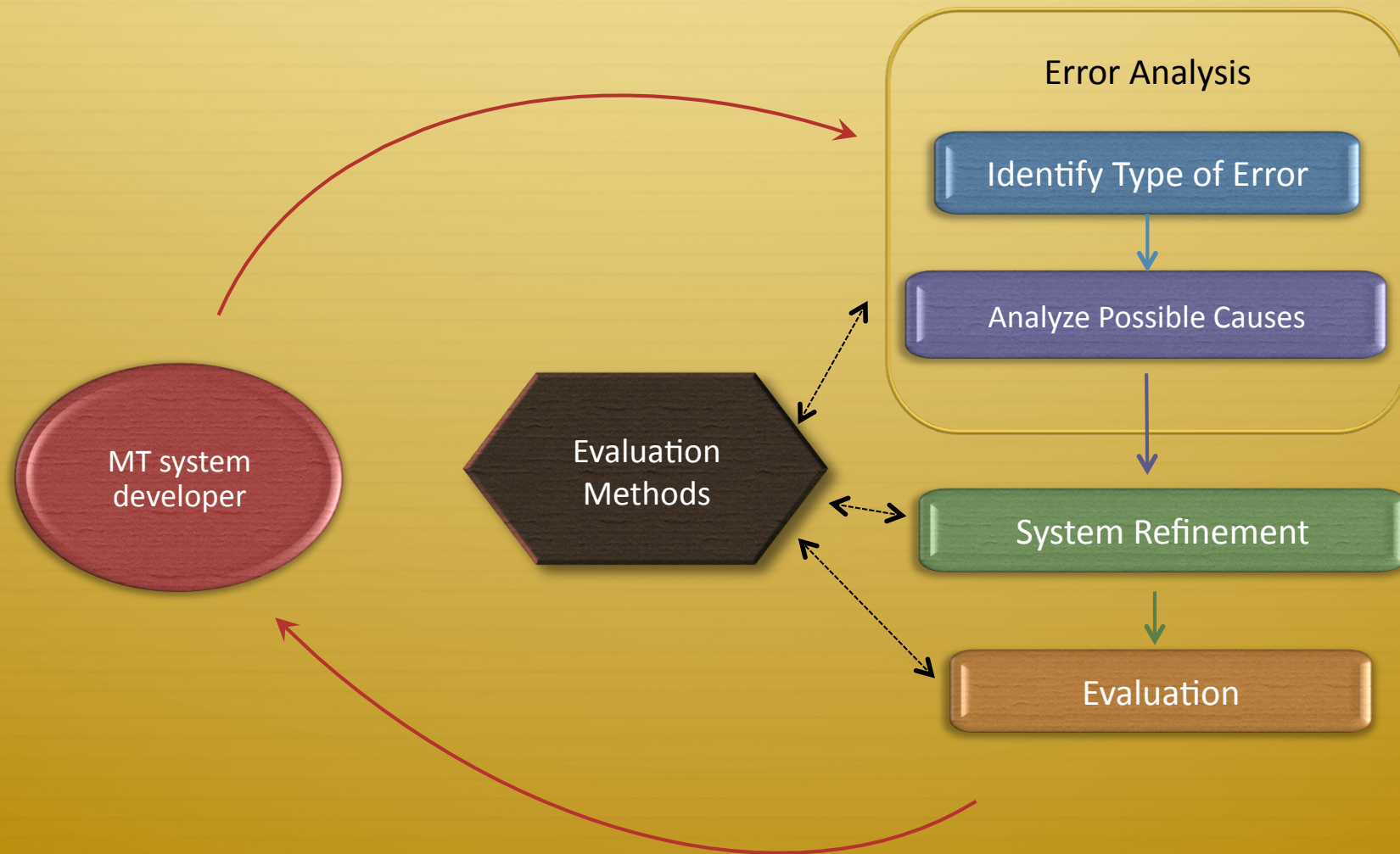




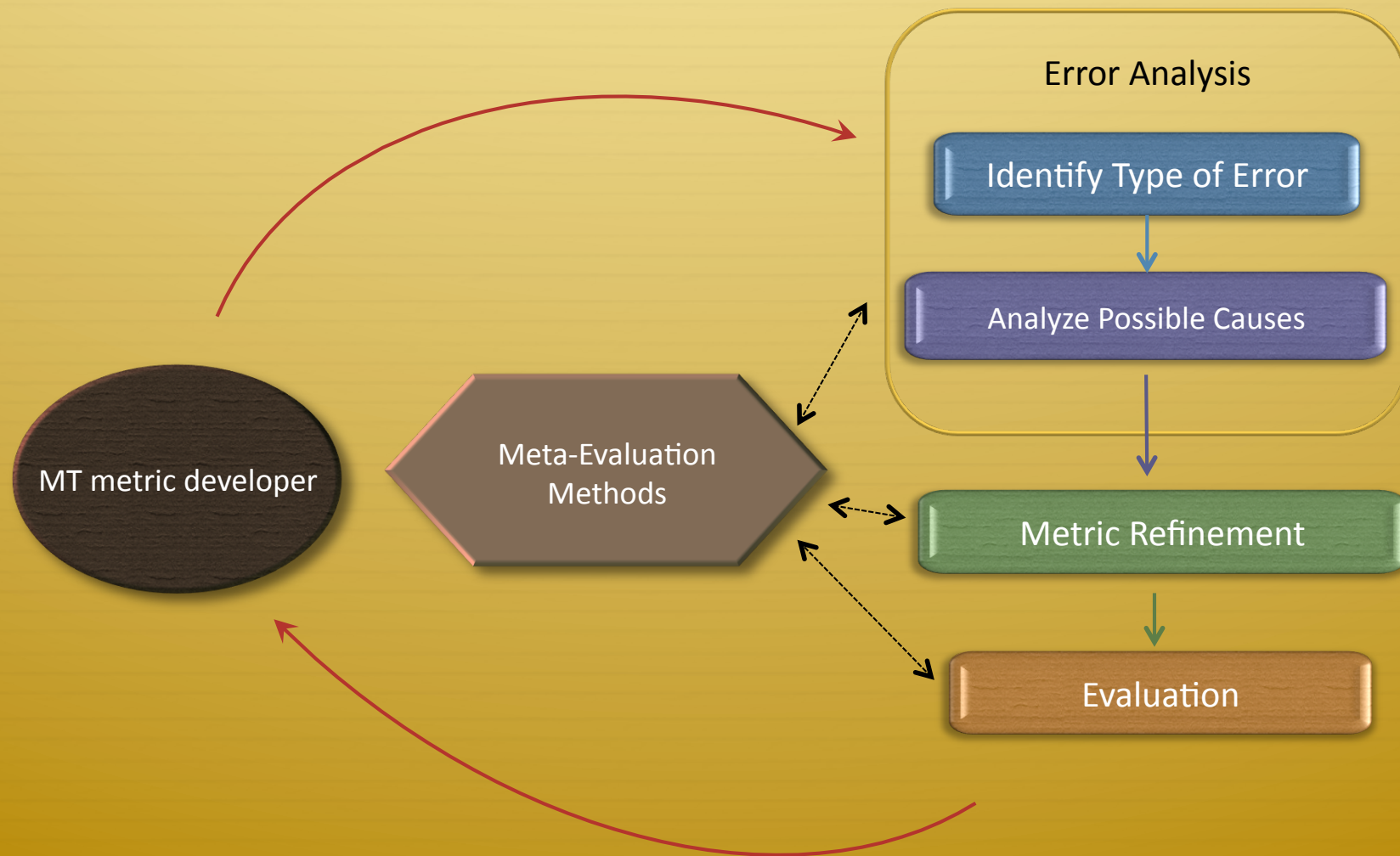
Asíya: An Open Toolkit for Automatic Machine Translation and (Meta-)Evaluation

Meritxell González
February 18th, 2013

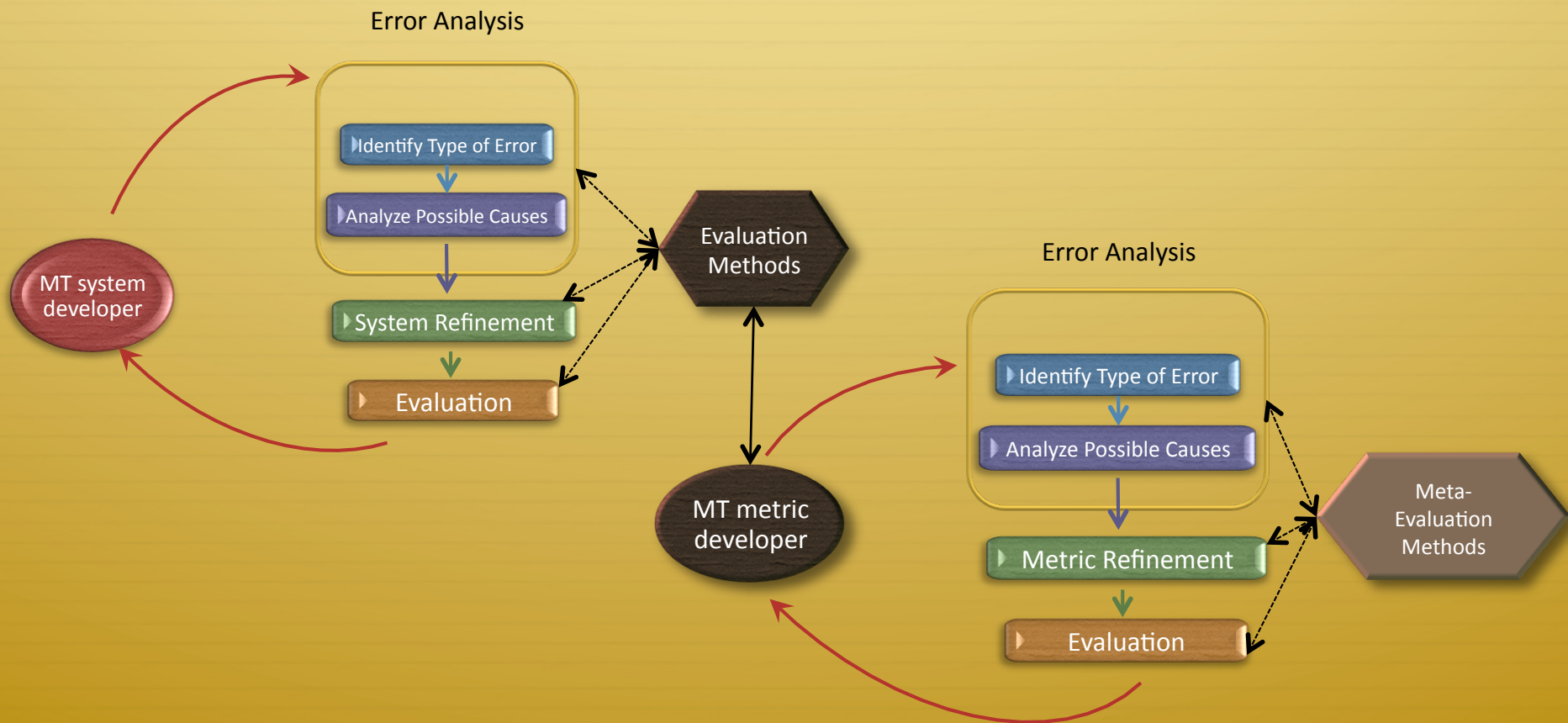
MT Development cycle



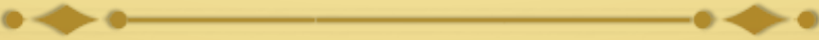
MT Development cycle



MT Development cycle



Metrics and Meta-metrics



- ✧ Languages:

- ✧ English, Spanish, Catalan

- ✧ Also: Arabic, Czech, French, German, Czech, Romanian

- ✧ Similarity principles

- ✧ Precision, recall, overlap

- ✧ Linguistic layers:

- ✧ Lexical, Syntactic, Semantic

- ✧ Confidence estimation

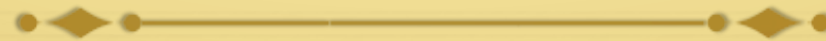
- ✦ 813 metrics are available for language 'en'

METRICS = 1-PER, TER, TERbase, TERp, TERp-A, WER, ALGP, ALGN, ALGNs, BLEU, BLEU-1, BLEU-2, BLEU-3, BLEU-4, BLEU-5, BLEU-6, CE-BiDictA, CE-BiDictO, CE-Nc, CE-Ne, CE-Oc, CE-Oc-CP, CE-Oc-CPpl, CE-ipplC, CE-ipplC-CE-length, CE-lognCP, CE-lognPL, CE-long, CE-ovv, CE-short, CE-scrippC, CE-scrippC-CE-srlen, CE-srlogcp, CE-srlogcp-CE-srcoov, CE-symbols, CP-Oc*, CP-Oc(ADJP), CP-Oc(ADVP), CP-Oc(CONJP), CP-Oc(FRAG), CP-Oc(INTJ), CP-Oc(LST), CP-Oc(NAC), CP-Oc(NP), CP-Oc(NX), CP-Oc(O), CP-Oc(PP), CP-Oc(PRN), CP-Oc(PRT), CP-Oc(QP), CP-Oc(RRC), CP-Oc(S), CP-Oc(SBAR), CP-Oc(SINV), CP-Oc(SO), CP-Oc(UCP), CP-Oc(VP), CP-Oc(WHADJP), CP-Oc(WHADVP), CP-Oc(WHNP), CP-Oc(WHPP), CP-Oc(X), CP-Oc(Y), CP-Oc(Z), CP-Oc("), CP-Oc(,), CP-Oc(,), CP-Oc(,), CP-Oc(,), CP-Oc(CC), CP-Oc(CD), CP-Oc(DT), CP-Oc(EX), CP-Oc(F), CP-Oc(FW), CP-Oc(IN), CP-Oc(J), CP-Oc(JJ), CP-Oc(JJS), CP-Oc(LS), CP-Oc(LS), CP-Oc(MD), CP-Oc(N), CP-Oc(NN), CP-Oc(NNP), CP-Oc(NNPS), CP-Oc(NNS), CP-Oc(P), CP-Oc(PDT), CP-Oc(POS), CP-Oc(PRPS), CP-Oc(PRP), CP-Oc(R), CP-Oc(RB), CP-Oc(RBR), CP-Oc(RBS), CP-Oc(RP), CP-Oc(SYM), CP-Oc(TO), CP-Oc(UH), CP-Oc(V), CP-Oc(VB), CP-Oc(VBD), CP-Oc(VBG), CP-Oc(VBN), CP-Oc(VBP), CP-Oc(VBZ), CP-Oc(W), CP-Oc(WDT), CP-Oc(WPS), CP-Oc(WP), CP-Oc(WRB), CP-Oc("), CP-STM-1, CP-STM-2, CP-STM-3, CP-STM-4, CP-STM-5, CP-STM-6, CP-STM-7, CP-STM-8, CP-STM-9, CP-STM-10, CP-STM-11, CP-STM-12, CP-STM-13, CP-STM-14, CP-STM-15, CP-STM-16, CP-STM-17, CP-STM-18, CP-STM-19, CP-STM-20, CP-STM-21, CP-STM-22, CP-STM-23, CP-STM-24, CP-STM-25, CP-STM-26, CP-STM-27, CP-STM-28, CP-STM-29, CP-STM-30, CP-STM-31, CP-STM-32, CP-STM-33, CP-STM-34, CP-STM-35, CP-STM-36, CP-STM-37, CP-STM-38, CP-STM-39, CP-STM-40, CP-STM-41, CP-STM-42, CP-STM-43, CP-STM-44, CP-STM-45, CP-STM-46, CP-STM-47, CP-STM-48, CP-STM-49, CP-STM-50, CP-STM-51, CP-STM-52, CP-STM-53, CP-STM-54, CP-STM-55, CP-STM-56, CP-STM-57, CP-STM-58, CP-STM-59, CP-STM-60, CP-STM-61, CP-STM-62, CP-STM-63, CP-STM-64, CP-STM-65, CP-STM-66, CP-STM-67, CP-STM-68, CP-STM-69, CP-STM-70, CP-STM-71, CP-STM-72, CP-STM-73, CP-STM-74, CP-STM-75, CP-STM-76, CP-STM-77, CP-STM-78, CP-STM-79, CP-STM-80, CP-STM-81, CP-STM-82, CP-STM-83, CP-STM-84, CP-STM-85, CP-STM-86, CP-STM-87, CP-STM-88, CP-STM-89, CP-STM-90, CP-STM-91, CP-STM-92, CP-STM-93, CP-STM-94, CP-STM-95, CP-STM-96, CP-STM-97, CP-STM-98, CP-STM-99, CP-STM-100, CP-STM-101, CP-STM-102, CP-STM-103, CP-STM-104, CP-STM-105, CP-STM-106, CP-STM-107, CP-STM-108, CP-STM-109, CP-STM-110, CP-STM-111, CP-STM-112, CP-STM-113, CP-STM-114, CP-STM-115, CP-STM-116, CP-STM-117, CP-STM-118, CP-STM-119, CP-STM-120, CP-STM-121, CP-STM-122, CP-STM-123, CP-STM-124, CP-STM-125, CP-STM-126, CP-STM-127, CP-STM-128, CP-STM-129, CP-STM-130, CP-STM-131, CP-STM-132, CP-STM-133, CP-STM-134, CP-STM-135, CP-STM-136, CP-STM-137, CP-STM-138, CP-STM-139, CP-STM-140, CP-STM-141, CP-STM-142, CP-STM-143, CP-STM-144, CP-STM-145, CP-STM-146, CP-STM-147, CP-STM-148, CP-STM-149, CP-STM-150, CP-STM-151, CP-STM-152, CP-STM-153, CP-STM-154, CP-STM-155, CP-STM-156, CP-STM-157, CP-STM-158, CP-STM-159, CP-STM-160, CP-STM-161, CP-STM-162, CP-STM-163, CP-STM-164, CP-STM-165, CP-STM-166, CP-STM-167, CP-STM-168, CP-STM-169, CP-STM-170, CP-STM-171, CP-STM-172, CP-STM-173, CP-STM-174, CP-STM-175, CP-STM-176, CP-STM-177, CP-STM-178, CP-STM-179, CP-STM-180, CP-STM-181, CP-STM-182, CP-STM-183, CP-STM-184, CP-STM-185, CP-STM-186, CP-STM-187, CP-STM-188, CP-STM-189, CP-STM-190, CP-STM-191, CP-STM-192, CP-STM-193, CP-STM-194, CP-STM-195, CP-STM-196, CP-STM-197, CP-STM-198, CP-STM-199, CP-STM-200, CP-STM-201, CP-STM-202, CP-STM-203, CP-STM-204, CP-STM-205, CP-STM-206, CP-STM-207, CP-STM-208, CP-STM-209, CP-STM-210, CP-STM-211, CP-STM-212, CP-STM-213, CP-STM-214, CP-STM-215, CP-STM-216, CP-STM-217, CP-STM-218, CP-STM-219, CP-STM-220, CP-STM-221, CP-STM-222, CP-STM-223, CP-STM-224, CP-STM-225, CP-STM-226, CP-STM-227, CP-STM-228, CP-STM-229, CP-STM-230, CP-STM-231, CP-STM-232, CP-STM-233, CP-STM-234, CP-STM-235, CP-STM-236, CP-STM-237, CP-STM-238, CP-STM-239, CP-STM-240, CP-STM-241, CP-STM-242, CP-STM-243, CP-STM-244, CP-STM-245, CP-STM-246, CP-STM-247, CP-STM-248, CP-STM-249, CP-STM-250, CP-STM-251, CP-STM-252, CP-STM-253, CP-STM-254, CP-STM-255, CP-STM-256, CP-STM-257, CP-STM-258, CP-STM-259, CP-STM-260, CP-STM-261, CP-STM-262, CP-STM-263, CP-STM-264, CP-STM-265, CP-STM-266, CP-STM-267, CP-STM-268, CP-STM-269, CP-STM-270, CP-STM-271, CP-STM-272, CP-STM-273, CP-STM-274, CP-STM-275, CP-STM-276, CP-STM-277, CP-STM-278, CP-STM-279, CP-STM-280, CP-STM-281, CP-STM-282, CP-STM-283, CP-STM-284, CP-STM-285, CP-STM-286, CP-STM-287, CP-STM-288, CP-STM-289, CP-STM-290, CP-STM-291, CP-STM-292, CP-STM-293, CP-STM-294, CP-STM-295, CP-STM-296, CP-STM-297, CP-STM-298, CP-STM-299, CP-STM-300, CP-STM-301, CP-STM-302, CP-STM-303, CP-STM-304, CP-STM-305, CP-STM-306, CP-STM-307, CP-STM-308, CP-STM-309, CP-STM-310, CP-STM-311, CP-STM-312, CP-STM-313, CP-STM-314, CP-STM-315, CP-STM-316, CP-STM-317, CP-STM-318, CP-STM-319, CP-STM-320, CP-STM-321, CP-STM-322, CP-STM-323, CP-STM-324, CP-STM-325, CP-STM-326, CP-STM-327, CP-STM-328, CP-STM-329, CP-STM-330, CP-STM-331, CP-STM-332, CP-STM-333, CP-STM-334, CP-STM-335, CP-STM-336, CP-STM-337, CP-STM-338, CP-STM-339, CP-STM-340, CP-STM-341, CP-STM-342, CP-STM-343, CP-STM-344, CP-STM-345, CP-STM-346, CP-STM-347, CP-STM-348, CP-STM-349, CP-STM-350, CP-STM-351, CP-STM-352, CP-STM-353, CP-STM-354, CP-STM-355, CP-STM-356, CP-STM-357, CP-STM-358, CP-STM-359, CP-STM-360, CP-STM-361, CP-STM-362, CP-STM-363, CP-STM-364, CP-STM-365, CP-STM-366, CP-STM-367, CP-STM-368, CP-STM-369, CP-STM-370, CP-STM-371, CP-STM-372, CP-STM-373, CP-STM-374, CP-STM-375, CP-STM-376, CP-STM-377, CP-STM-378, CP-STM-379, CP-STM-380, CP-STM-381, CP-STM-382, CP-STM-383, CP-STM-384, CP-STM-385, CP-STM-386, CP-STM-387, CP-STM-388, CP-STM-389, CP-STM-390, CP-STM-391, CP-STM-392, CP-STM-393, CP-STM-394, CP-STM-395, CP-STM-396, CP-STM-397, CP-STM-398, CP-STM-399, CP-STM-400, CP-STM-401, CP-STM-402, CP-STM-403, CP-STM-404, CP-STM-405, CP-STM-406, CP-STM-407, CP-STM-408, CP-STM-409, CP-STM-410, CP-STM-411, CP-STM-412, CP-STM-413, CP-STM-414, CP-STM-415, CP-STM-416, CP-STM-417, CP-STM-418, CP-STM-419, CP-STM-420, CP-STM-421, CP-STM-422, CP-STM-423, CP-STM-424, CP-STM-425, CP-STM-426, CP-STM-427, CP-STM-428, CP-STM-429, CP-STM-430, CP-STM-431, CP-STM-432, CP-STM-433, CP-STM-434, CP-STM-435, CP-STM-436, CP-STM-437, CP-STM-438, CP-STM-439, CP-STM-440, CP-STM-441, CP-STM-442, CP-STM-443, CP-STM-444, CP-STM-445, CP-STM-446,

Example

- ✦ El capitán descarta que el técnico abandone el banquillo del Barça por problemas con algunos de sus jugadores.
- ✦ The captain rejects that the coach leaves the Barça bench due problems with some of the players.
- ✦ The captain **descarta** that the technician abandon the **banquillo** of the Barça by problems with some of his players.
- ✦ The captain **discards** that the **technician** leaves **the bench of the Barça** by problems with some of his players.
- ✦ The captain **dismisses** the **technician** leaves the Barca bench due to problems with some of **their** players.
- ✦ The **master ruled** that the technician leaves the Barca bench by problems with some of his players.
- ✦ The captain rejects that the technician leaves the bench of the Barça for problems with some **of his(her,your) players**.
- ✦ The captain discards that the technician leaves the bench of the Barça **by problems** with some of his players.

Example - Overlap



✦ **The captain** rejects **that the** coach leaves **the Barça** bench due **problems with some of the players** .

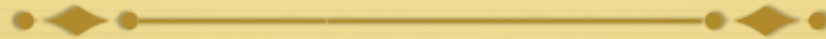
✦ { 'the'= 4; 'captain'= 1; 'coach'= 1; 'rejects'= 1; 'due'= 1; 'leaves'= 1; '**of**'= 1; 'players'= 1; 'that'= 1; 'barça'= 1; '.'= 1; 'bench'= 1; 'some'= 1; 'with'= 1; 'problems'= 1 } (15 different tokens)

✦ **The captain** descarta **that the** technician abandon the banquillo of **the Barça** by **problems with some of his players** .

✦ { '**the**' = 4; '**captain**'= 1; 'technician'= 1; 'descarta'= 1; 'of'= 2; '**players**'= 1; '**that**'= 1; '.'= 1; 'his'= 1; 'banquillo'= 1; '**problems**'= 1; 'abandon'= 1; 'by'= 1; '**barça**'= 1; '**some**'= 1; '**with**'= 1 } (16 different tokens)

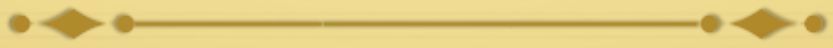
✦ O1 => hits: 13 and total: 25 = 13/25 = 0,52

Example - Overlap



- ✦ **The captain** rejects **that the** coach leaves **the Barça** bench due **problems with some of the players** .
- ✦ { 'the'= 4; 'captain'= 1; 'coach'= 1; 'rejects'= 1; 'due'= 1; 'leaves'= 1; 'of'= 1; 'players'= 1; 'that'= 1; 'barça'= 1; '.'= 1; 'bench'= 1; 'some'= 1; 'with'= 1; 'problems'= 1 } (15 different tokens)
- ✦ **The captain** descarta **that the** technician abandon the banquillo of **the Barça** by **problems with some of his players** .
- ✦ { 'the'= 4; 'captain'= 1; 'technician'= 1; 'descarta'= 1; 'of'= 2; 'players'= 1; 'that'= 1; '.'= 1; 'his'= 1; 'banquillo'= 1; 'problems'= 1; 'abandon'= 1; 'by'= 1; 'barça'= 1; 'some'= 1; 'with'= 1 } (16 different tokens)
- ✦ OI => hits: 13 and total: **25** = $13/25 = 0,52$

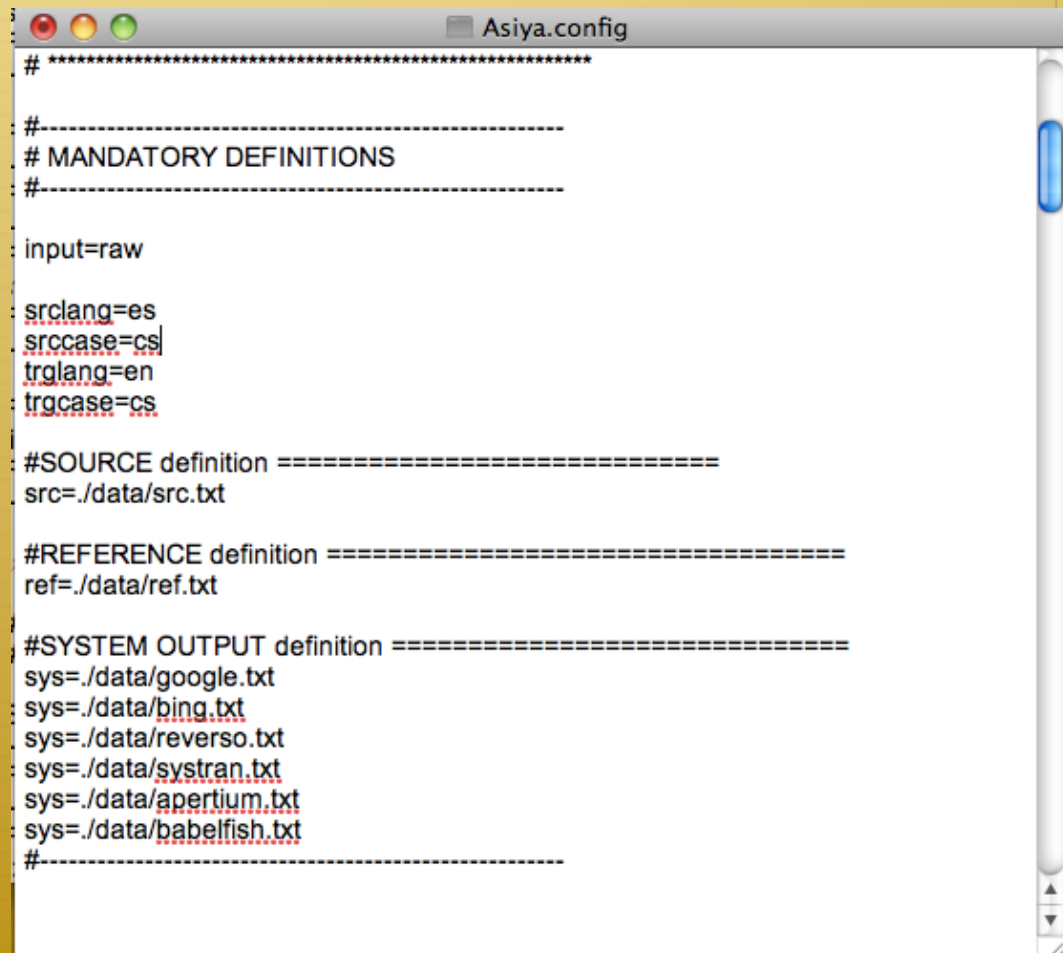
The Test Suit



- ✧ Asiya operates over test suites (or test beds).
- ✧ a test suite is a collection of test cases:
 - ✧ Source segment
 - ✧ Candidate translation(s)
 - ✧ Reference translation(s)

The Test Suit

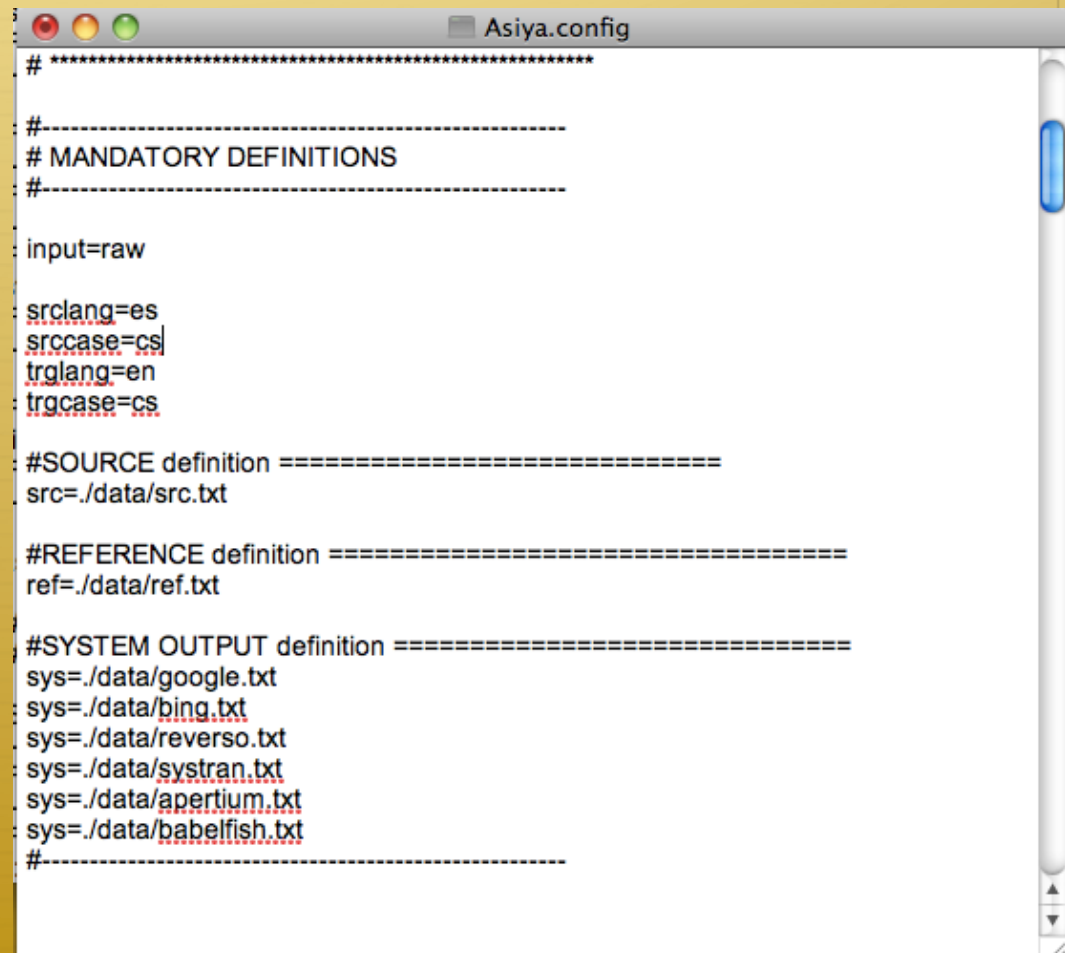
- ✧ Asiya.pl Asiya.config
- ✧ Asiya.config:

A screenshot of a text editor window titled 'Asiya.config'. The window shows a configuration file with various settings. The text is as follows:

```
# *****  
#-----  
# MANDATORY DEFINITIONS  
#-----  
  
input=raw  
  
srclang=es  
srccase=cs  
trglang=en  
trgcase=cs  
  
#SOURCE definition =====  
src=./data/src.txt  
  
#REFERENCE definition =====  
ref=./data/ref.txt  
  
#SYSTEM OUTPUT definition =====  
sys=./data/google.txt  
sys=./data/bing.txt  
sys=./data/reverso.txt  
sys=./data/systran.txt  
sys=./data/apertium.txt  
sys=./data/babelfish.txt  
#-----
```

General Options

- ✧ Input format
 - ✧ Raw
 - ✧ Nist
- ✧ Language pair
 - ✧ Srclang
 - ✧ Trglang
- ✧ Predefined sets of metrics, systems and references



The screenshot shows a text editor window titled "Asiya.config". The file contains configuration options for the Asiya system, organized into sections separated by dashed lines. The options include input format, language pairs, source and reference definitions, and system output definitions.

```
# *****  
#-----  
# MANDATORY DEFINITIONS  
#-----  
  
input=raw  
  
srclang=es  
srccase=cs  
trglang=en  
trgcase=cs  
  
#SOURCE definition =====  
src=./data/src.txt  
  
#REFERENCE definition =====  
ref=./data/ref.txt  
  
#SYSTEM OUTPUT definition =====  
sys=./data/google.txt  
sys=./data/bing.txt  
sys=./data/reverso.txt  
sys=./data/systran.txt  
sys=./data/apertium.txt  
sys=./data/babelfish.txt  
#-----
```

Eval

- ✧ Eval <schema>

- ✧ Single

- ✧ ULC

- ✧ Queen [AGPV05]

- ✧ Meta-Eval

- ✧ Learn

- ✧ Output format

- ✧ Metric matrix

- ✧ System matrix

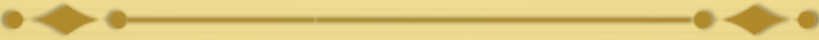
- ✧ Nist

- ✧ Granularity

- ✧ System, document,
segment

- ✧ Pdf, tex

Meta-Eval

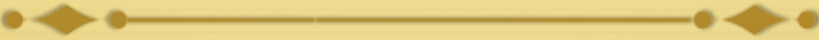


- ✧ Eval
- ✧ Meta-Eval <schemas> <criteria>
 - ✧ Correlation with assessments
 - ✧ Pearson
 - ✧ Spearman
 - ✧ Kendall tau
 - ✧ Orange [LO04]
 - ✧ King [AGPV05]
 - ✧ Consistency
- ✧ Learn

Meta-Eval

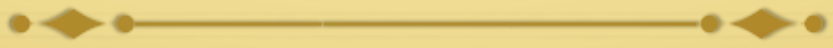
- ✧ Eval
- ✧ Meta-Eval <schemas> <criteria>
-ci <method> Asiya.config
 - ✧ Fisher [Fis24]
 - ✧ Bootstrap resampling [ET86]
 - ✧ Paired bootstrap resampling [Koe04]
Orange [LO04]
 - ✧ Options:
 - ✧ significance level
 - ✧ Asiya.pl -v -optimize <schemes>
<criteria> number of
resamplings
- ✧ Learn

Learning



- ✧ Eval
 - ✧ Asiya.pl -eval single -model <s>
- ✧ Meta-Eval
- ✧ Learn <scheme>
 - ✧ Perceptron
 - ✧ model <s>

Asiya Interfaces



- ✧ Asiya Online Interface

- ✧ A graphical interface to access an on-line version of Asiya.

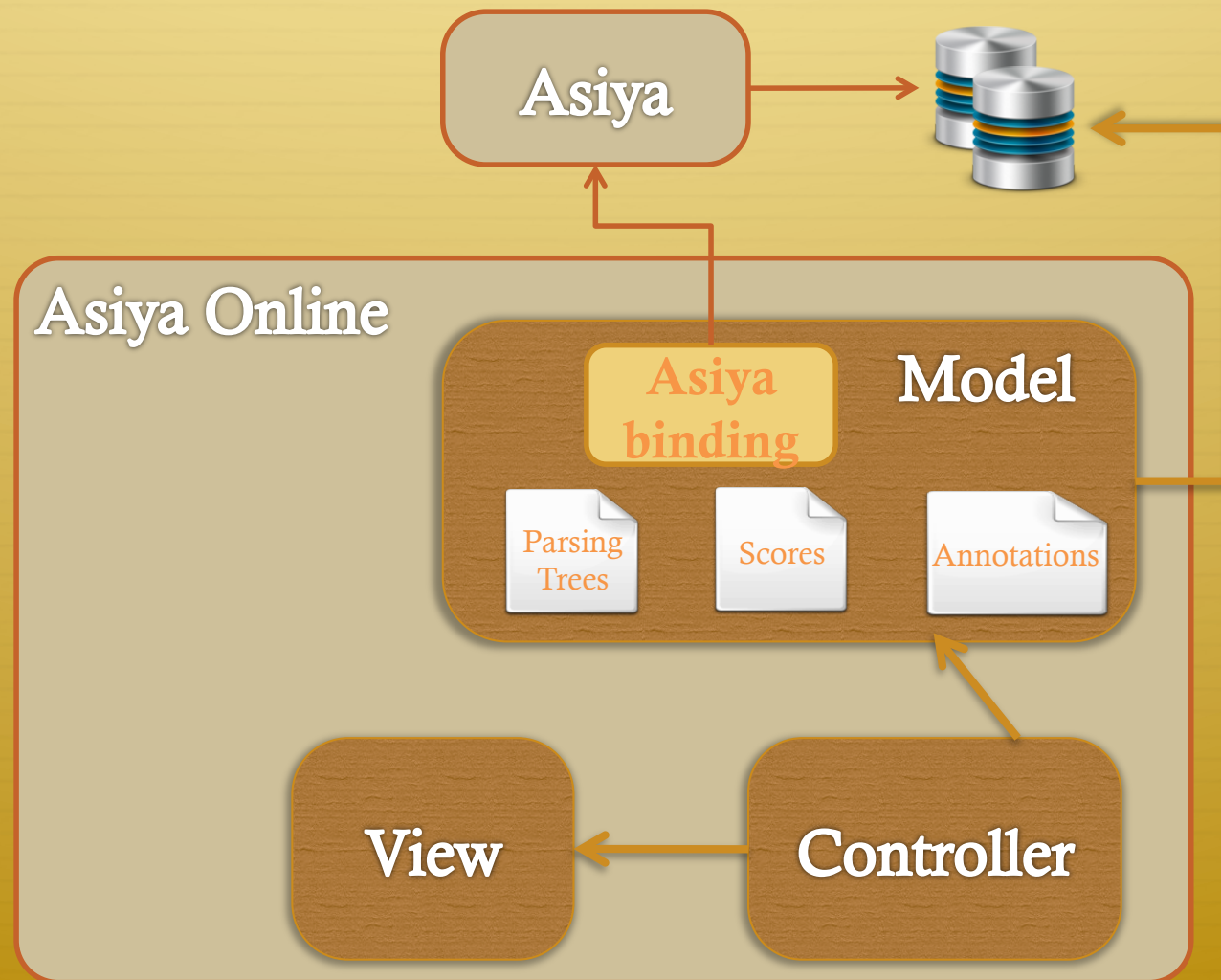
- ✧ Asiya tSearch

- ✧ online interface that allows to search for output translations (of a given testbed) that match some specific criteria related to their quality (as assessed by the automatic scores).

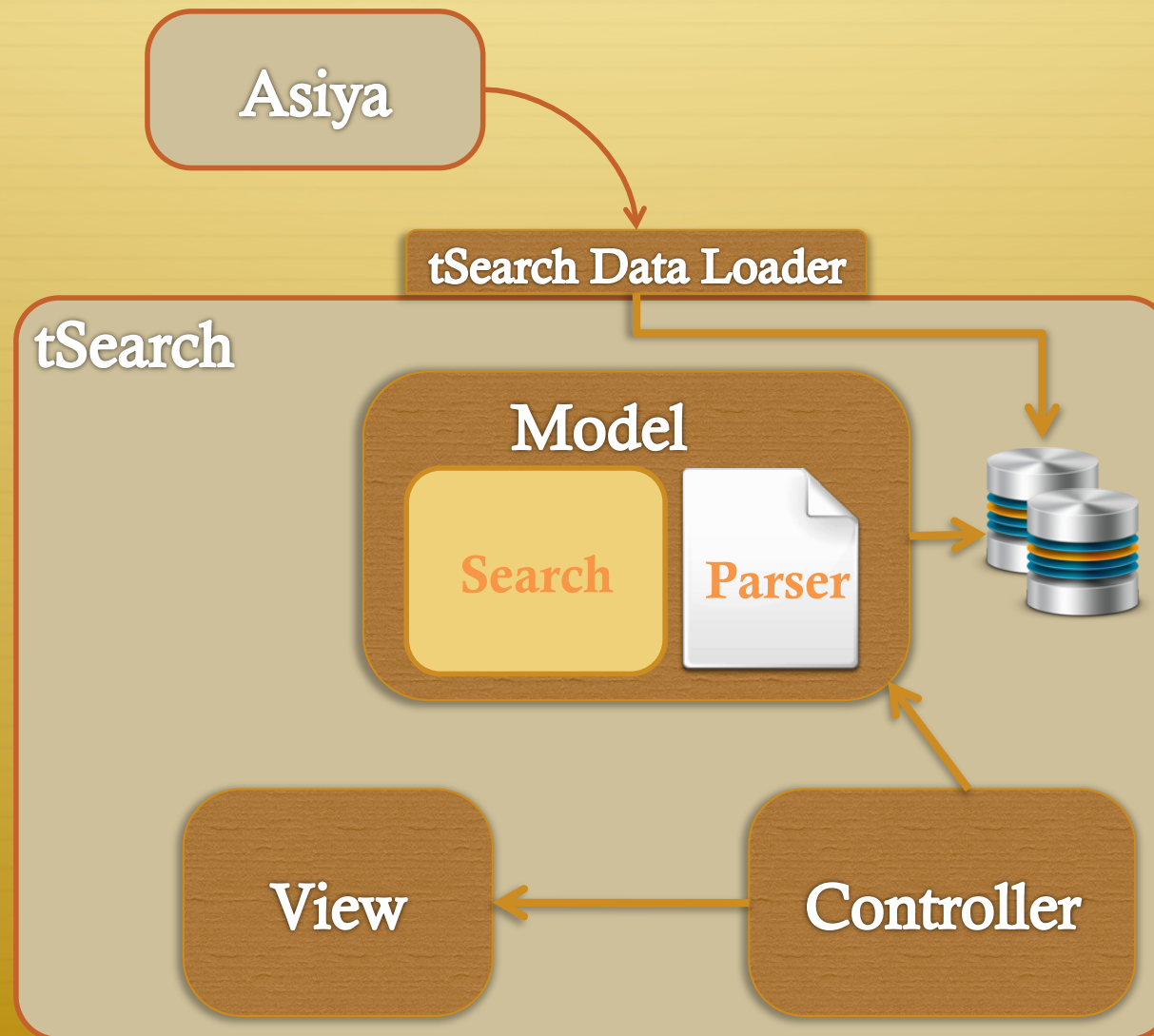
- ✧ Asiya Web Service

- ✧ A RESTful web service to access the Asiya evaluation.

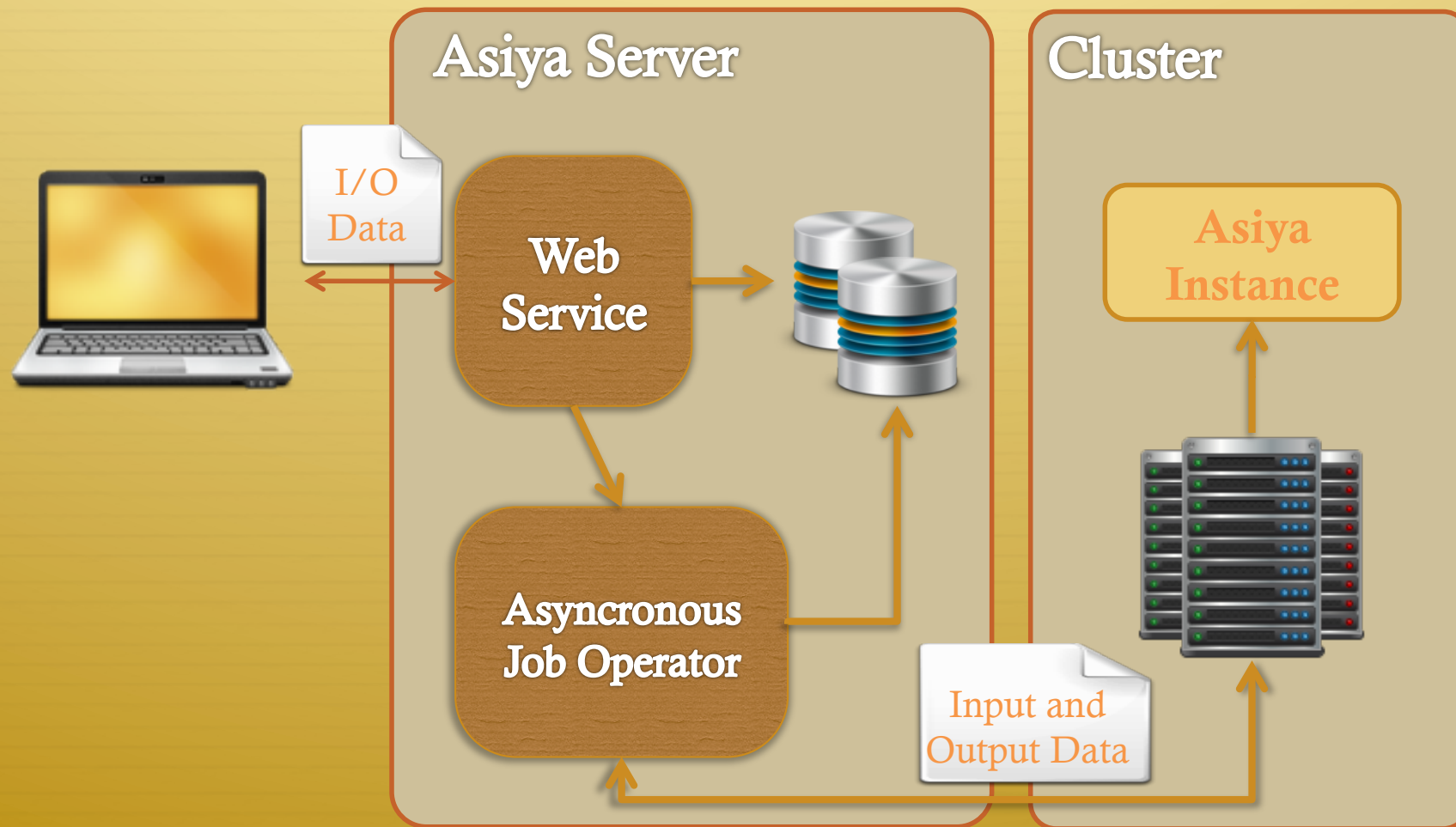
Asiya Online Interface



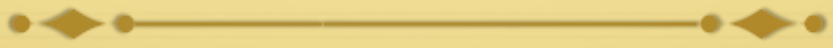
Asiya tSearch



Asiya Web Service

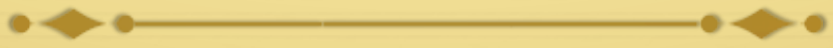


References



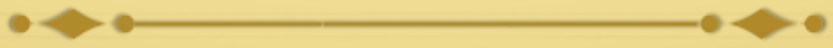
- ✦ Enrique Amigó, Julio Gonzalo, Anselmo Penas, and Felisa Verdejo. QARLA: a Framework for the Evaluation of Automatic Summarization. In Proceedings of the 43rd Annual Meeting of the Association for Computational Linguistics (ACL), pages 280–289, 2005.
- ✦ Chris Callison-Burch, Philipp Koehn, Christof Monz, and Josh Schroeder. Findings of the 2009 Workshop on Statistical Machine Translation. In Proceedings of the Fourth Workshop on Statistical Machine Translation, pages 1–28, 2009.
- ✦ Bradley Efron and Robert Tibshirani. Bootstrap Methods for Standard Errors, Confidence Intervals, and Other Measures of Statistical Accuracy. *Statistical Science*, 1(1):54–77, 1986.

References



- ✧ R. A. Fisher. On a Distribution Yielding the Error Functions of Several Well Known Statistics. In Proceedings of the International Congress of Mathematics, volume 2, pages 805–813, 1924.
- ✧ Jesús Giménez and Lluís Màrquez. Linguistic Features for Automatic MT Evaluation. To Appear in Machine Translation, 2010.
- ✧ Maurice Kendall. Rank Correlation Methods. Hafner Publishing Co, 1955.
- ✧ Philipp Koehn. Statistical Significance Tests for Machine Translation Evaluation. In Proceedings of the Conference on Empirical Methods in Natural Language Processing (EMNLP), pages 388–395, 2004.

References



- ✧ Chin-Yew Lin and Franz Josef Och. ORANGE: a Method for Evaluating Automatic Evaluation Metrics for Machine Translation. In Proceedings of the 20th International Conference on Computational Linguistics (COLING), pages 501–507, 2004.
- ✧ Karl Pearson. The life, letters and labours of Francis Galton. 1914. (3 volumes: 1914, 1924, 1930).
- ✧ Charles Spearman. The Proof and Measurement of Association Between Two Rings. American Journal of Psychology, 15:72–101, 1904.