



WORD SENSE DISAMBIGUATION TOOL

(CODENAME: SEMANTIC PROCESSOR)

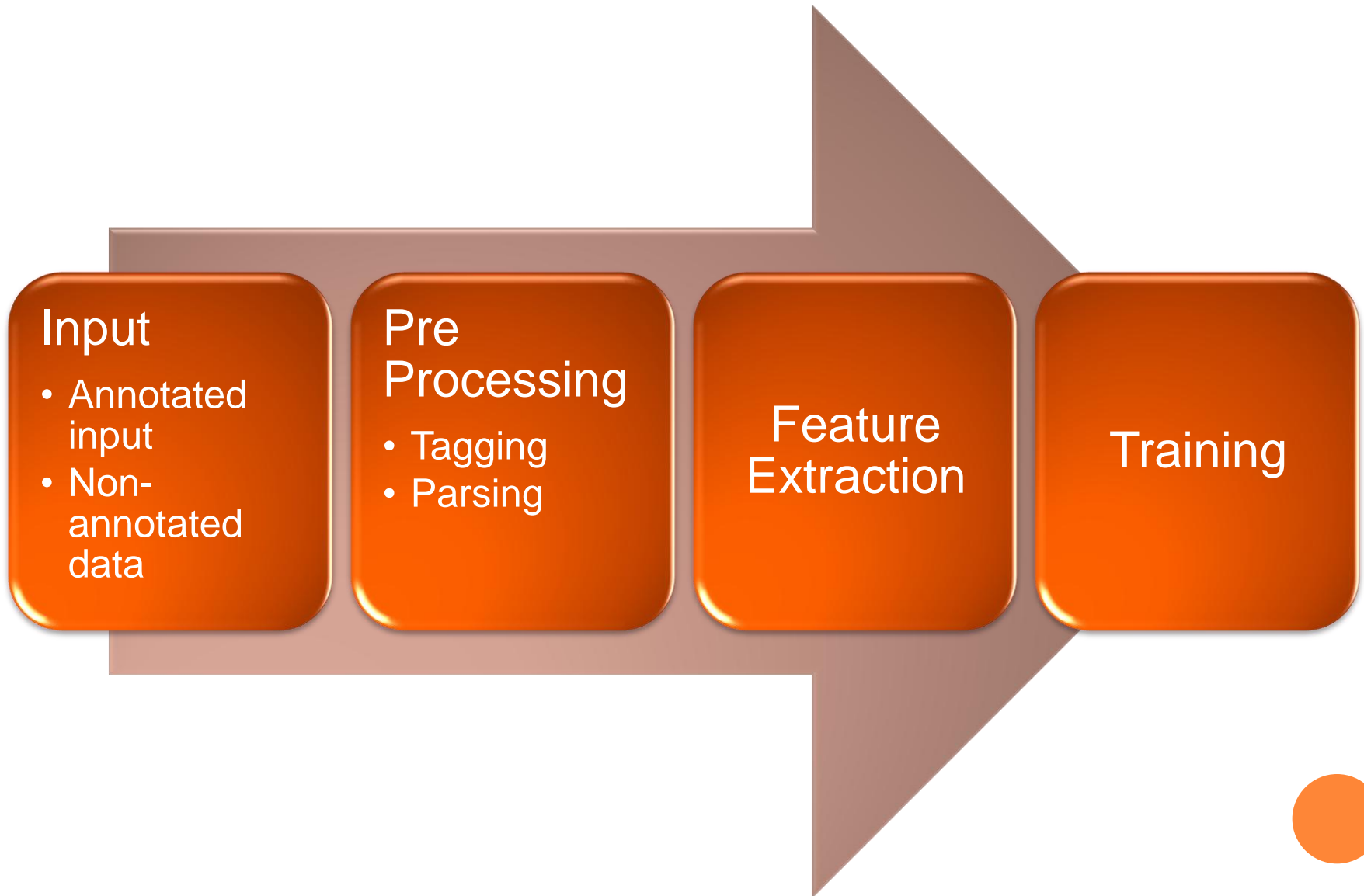
A project by Marcel Köster & Jonas Sunde

GENERAL INFORMATION

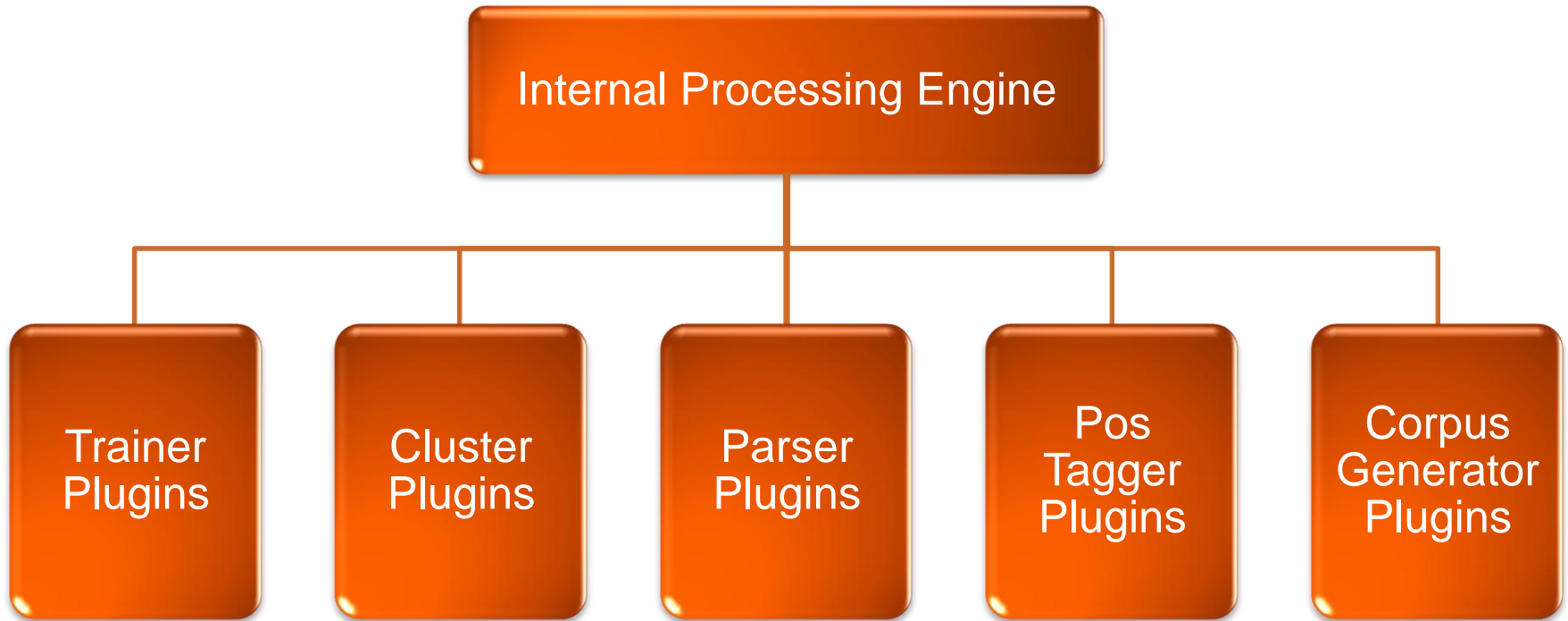
- Developed in winter term 2008/2009
- Usage: disambiguation of words
- Supervised “self learning” system
(...using a custom maximum entropy algorithm)
- Implementation: plugin based approach



INTERNAL TRANSFORMATION PROCESS

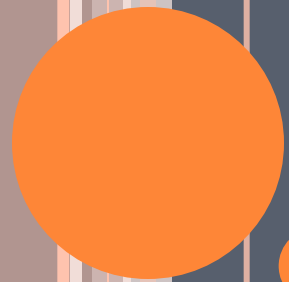


EXTENSIBILITY & PLUGINS



INTERNAL PROCESSES (GUI)





FEATURES (CLUSTER PLUGINS)



FEATURES (CLUSTER PLUGINS)

Word Range Clusterer

Ich melde mich im Unterricht.

- Extracts a word range with a given length relative to the index of the word (word: “melde”, index: 1)
- Several implementations:
 - Word Range Clusterer
-> Ich | mich (word range: 1)
 - PosTag Word Range Clusterer
-> PPER | PRF (word range: 1)



FEATURES (CLUSTER PLUGINS)

Clause Clusterer

Ich melde mich im Unterricht.

Ich	melde	mich	im	Unterricht	.
SUBJ	ROOT	OBJA	PP	PN	PUNCT
PPER	VVFIN	PRF	APPART	NN	\$SBI

- Extracts words or pos tags for given dependencies
- Sample:
Dependency: SUBJ
 - Word: Ich
 - PosTag: PPER



FEATURES (CLUSTER PLUGINS)

GermaNet Clusterer

Ich melde mich im Unterricht.

Ich	melde	mich	im	Unterricht	.
SUBJ	ROOT	OBJA	PP	PN	PUNCT
PPER	VVFIN	PRF	APPART	NN	\$SBI

- Uses the GermaNet API to resolve hyponymies, meronymies, ... with a maximum depth for specified parts of the sentence
- Two implementations (Depth 2):
 - Dependency Clusterer:
(PN, Hypernymy, 2) -> Bildungsveranstaltung, Veranstaltung
 - Pos Tag Clusterer:
(NN, Hypernymy, 2) -> Bildungsveranstaltung, Veranstaltung



FEATURES (CLUSTER PLUGINS)

GermaNet Superordinate PosTag Clusterer

Ich melde mich im Unterricht.

Ich	melde	mich
PPER	VVFIN	PRF
PRO	V	PRO

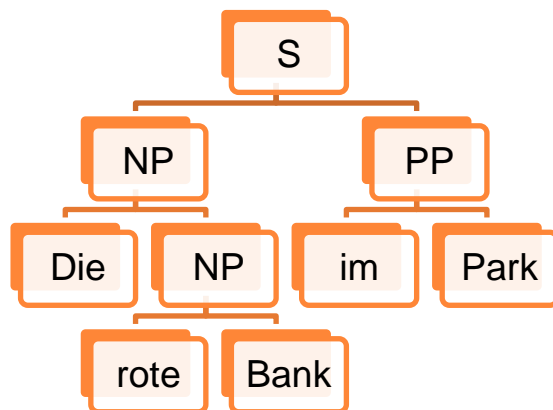
- Uses the GermaNet API to resolve hyponymies, meronymies, ... with a maximum depth for specified parts of the sentence with the same superordinate postags
 - (PRO, Hypernymy, 2):
 - (PPER, Hypernymy, 2) -> geistige_Wahrnehmung, Seele
 - (PRF, Hypernymy, 2) -> Sein, Seele



FEATURES (CLUSTER PLUGINS)

Sentence Phrase Clusterer

Sentence Phrase Pos Tag Clusterer



- Selects a subtree labeled with a phrase name (e.g. NP) from the parsed tree
- Uses the words within the subtree
- Example: NP
-> rote | Bank
- Selects a subtree labeled with a phrase name from the parsed tree
- Uses the pos tags within the subtree
- Example:
-> ADJA | NN



The left side of the slide features a series of vertical stripes in shades of brown, tan, and white. To the right of these stripes are several orange circles of varying sizes, arranged in a vertical line that tapers towards the bottom.

QUESTIONS ?

The left side of the slide features a series of vertical stripes in shades of brown, tan, and white. To the right of these stripes are several orange circles of varying sizes, arranged in a vertical line that tapers towards the bottom.

THANK YOU