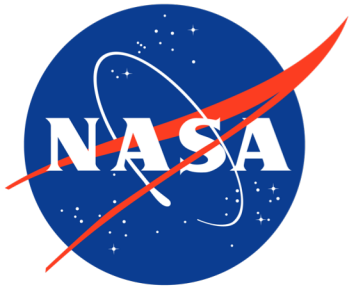
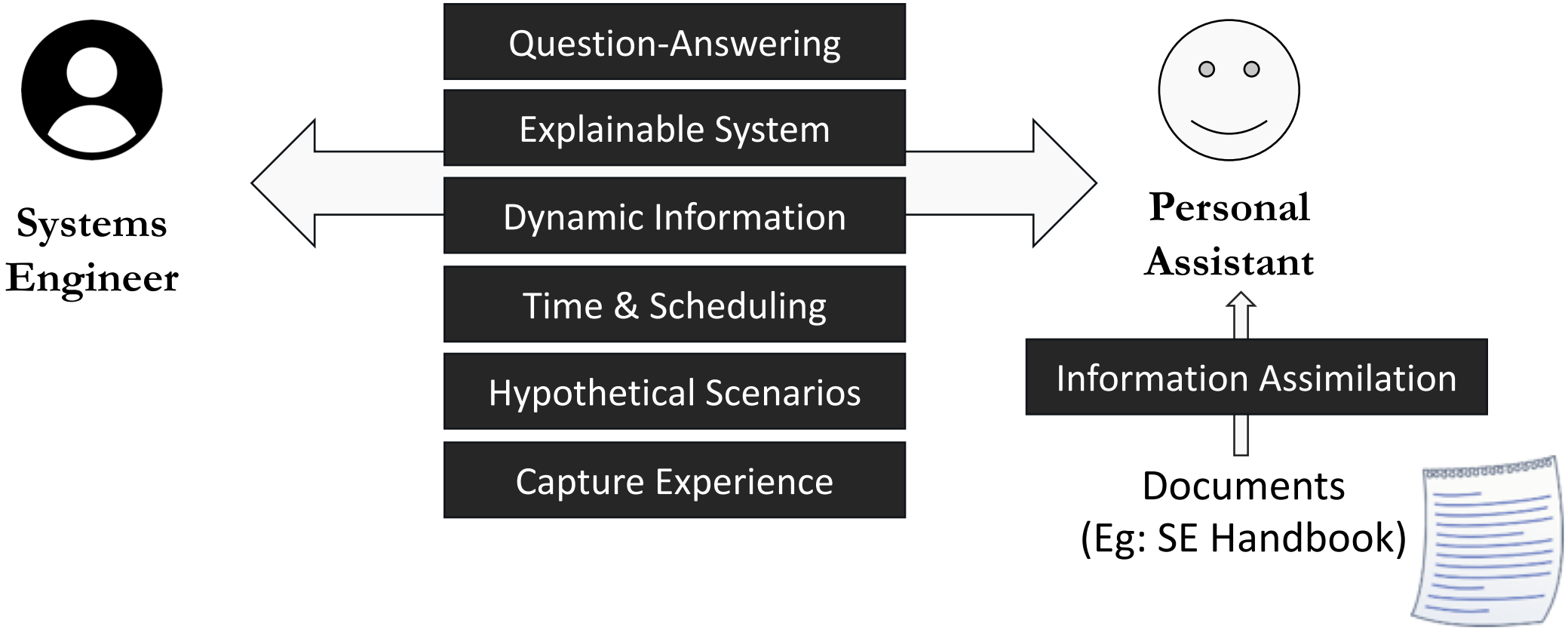


Open Information Extractor for SEVA

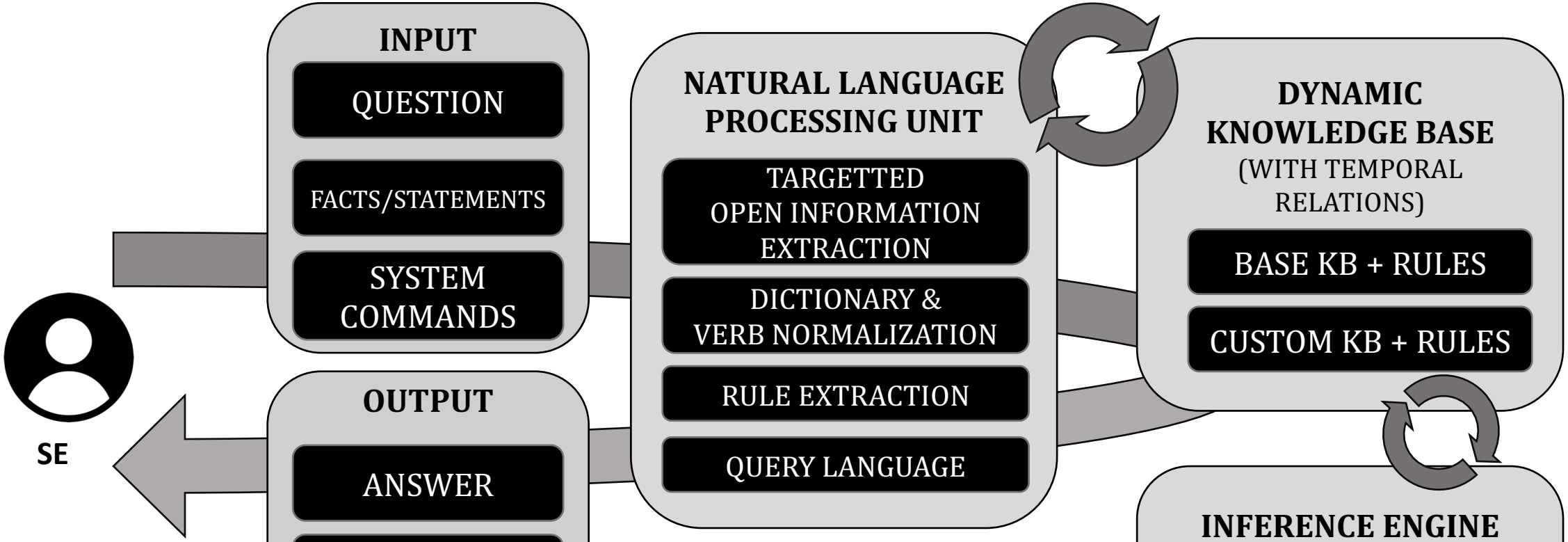
Jitin Krishnan
Mentor: Patrick Coronado

Other Project Members:
Trevor Reed (NASA JPL)
Ben Brumback (George Mason University)
Youngyo Na (Rowan University)

SEVA: A Systems Engineer's Virtual Assistant



SEVA: ARCHITECTURAL COMPONENTS

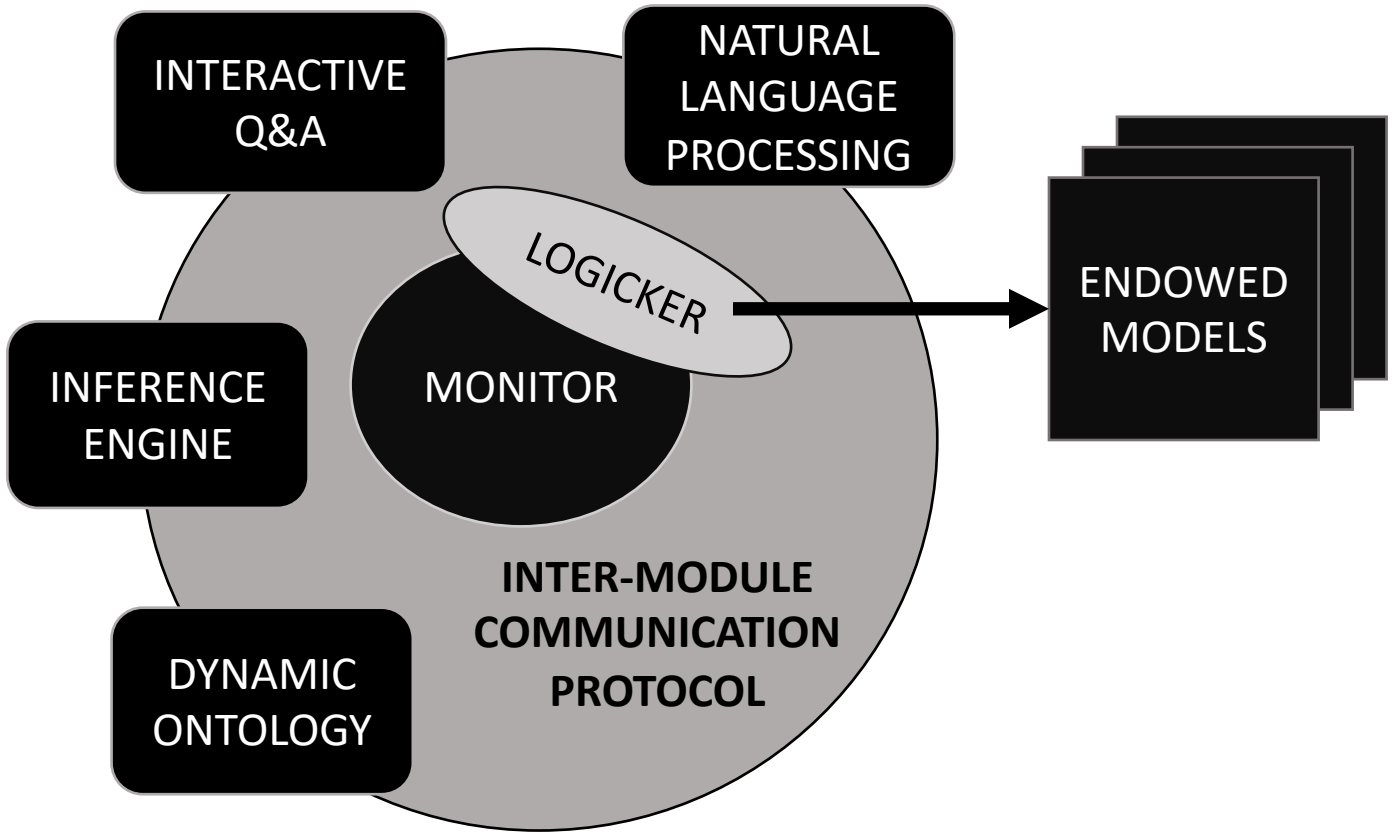


*in conversational mode (rule/experience building)

Motivation

- Information Assimilation
- Explainability
- Personal System
- Domain-specific Common Sense

SEVA: SYSTEM CONCEPT



SEVA: Knowledge Representation Example

Sentence: Instruments have thermal zones

Subject — Verb/Predicate — Object

NLP

First Order Logic (FOL) Formalism:

$$\forall x \quad \exists y \quad \texttt{Instrument}(x) \wedge \texttt{Thermal Zone}(y) \Rightarrow \texttt{have}(x,y)$$

Concepts: **Instrument, Thermal Zone** (like classes in OOP)

Instances: **x, y**

Relationship: **have**

KB

∀ = for all
∃ = there exists

SEVA: Knowledge Representation Example

ABox

Instrument(STI)
STI is an instrument

part of(MassSpectrometerAP8717, STI)
MassSpectrometerAP8717 is a part of STI

Instantiations and relation between instances

TBox

$\boldsymbol{Conduit \equiv Pipe}$
Conduit is same as Pipe

$\boldsymbol{MassSpectrometer \sqsubseteq Spectrometer}$
Subclass relationship

Concepts and relations

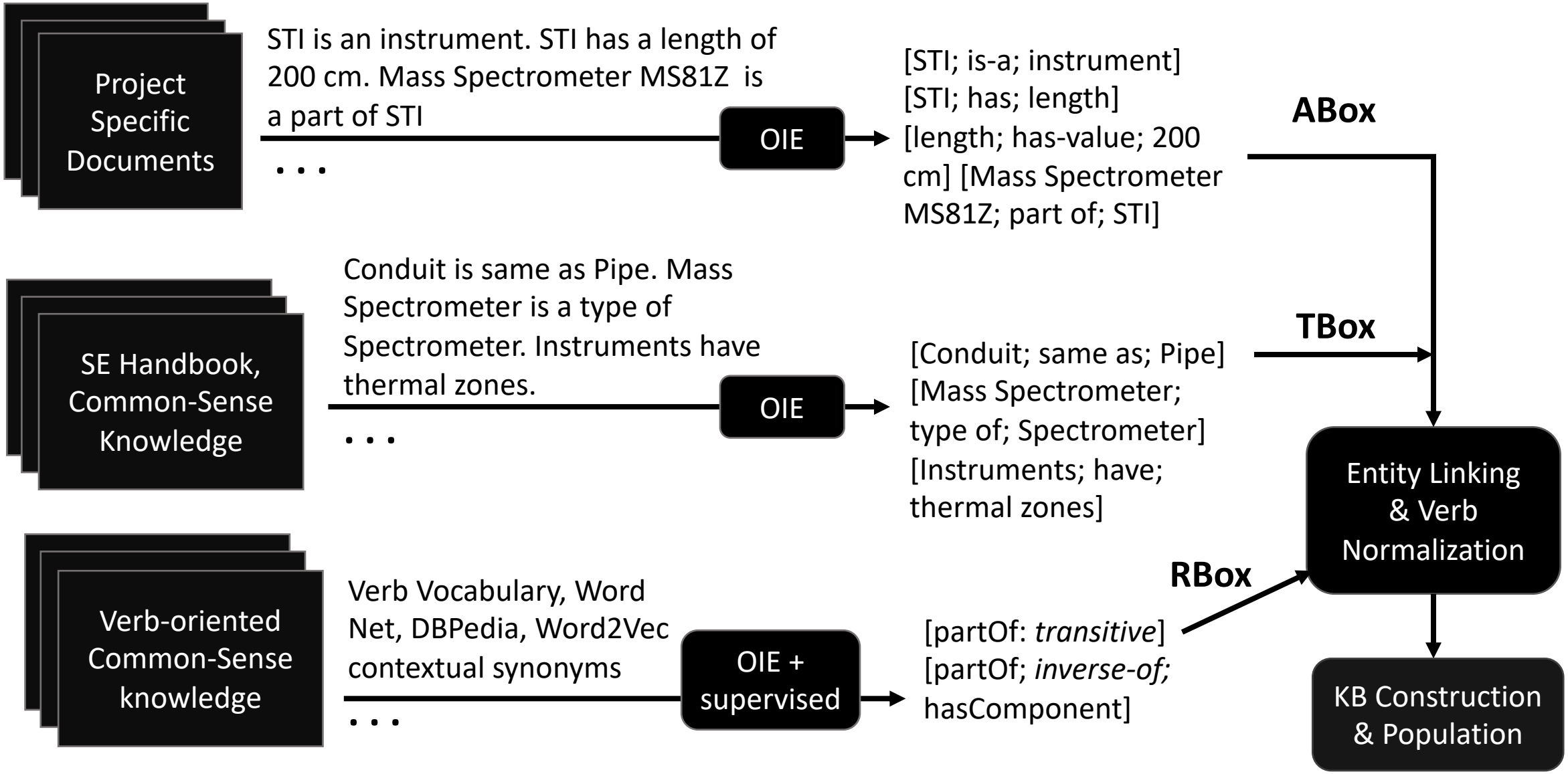
RBox

$\boldsymbol{partOf \circ partOf \sqsubseteq partOf}$
Transitive property of the role

$\boldsymbol{partOf \equiv hasComponent^-}$
Inverse property of two roles

Relations between relations

SEVA: Types of Knowledge



Open Information Extraction

- Identifies wide range of domain-independent relations
- Traditional Information Extraction: uses predetermined templates

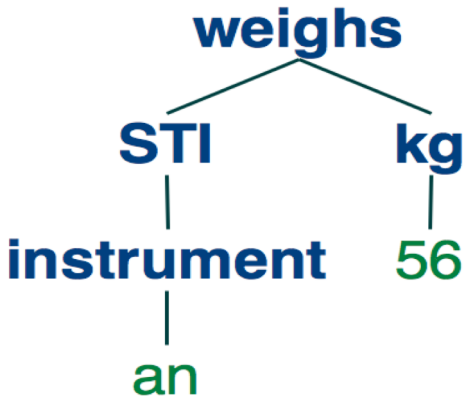
Examples: `Stanford Open IE`, `Open IE by AI2`, `ClausIE`

- For the experiment the domain language complexity is reduced – we work with only **simple English** sentences

NLP Basics

Sentence:

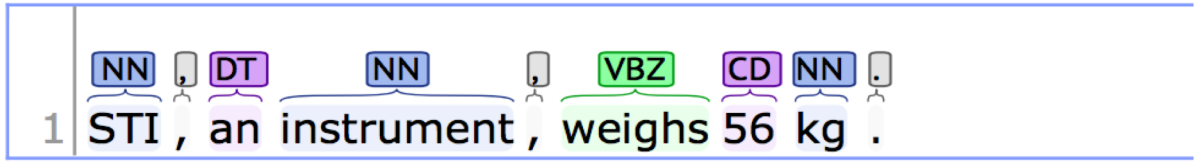
`STI, an instrument, weighs 56 kg`



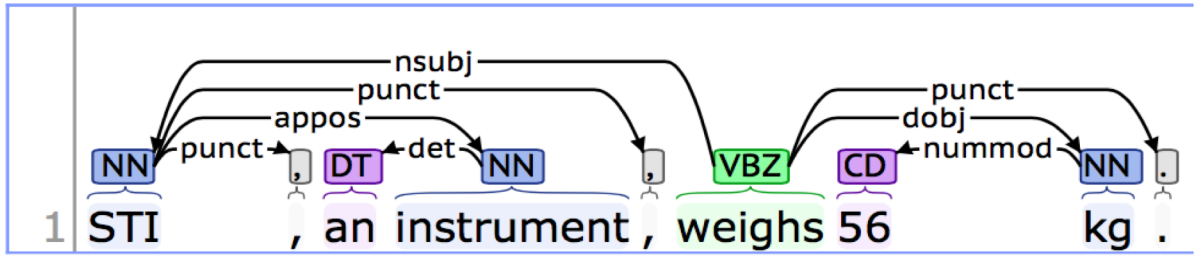
Dependency Parse Tree

Part-of-Speech:

NN = Noun
VBZ = Verb form



Basic Dependencies:



SEVA – Targeted Open IE (TOIE)

Pattern Matching on the dependency tree

Extracting relations of type:
`is-a`, **`transitive-verb`**, **`has-property`**, **`has-value`**

A set of universal dependencies:
`nsubj`, **`dobj`**, **`case`**, **`nmod`**, **`compound`**, **`amod`**

Chunking example:
`"NP:{(<CD.*>|<JJ.*>)<NN.*>+}"`

NLTK, Stanford CoreNLP, POS Tagger

SEVA-TOIE: Comparison

Input Sentence	Stanford Open IE	ClausIE	AI2 Open IE	SEVA-TOIE
STI, an instrument, has a 2500 pixel CCD detector	("STI" "has" "2500 pixel CCD detector") *incomplete/ missing information*: "STI is an instrument"	("STI" "is" "an instrument") ("STI" "has" "a 2500 pixel CCD detector")	(STI; has; a 2500 pixel CCD detector) (STI; [is]; an instrument)	(STI; has; CCD detector) (STI; is-a; instrument) (CCD detector; has-property; 2500 pixel)
STI is an instrument with a TRL value of 5	("STI" "is" "instrument") ("STI" "is instrument with" "TRL value of 5") ("instrument" "is with" "TRL value of 5") ("STI" "is instrument with" "TRL value")	("STI" "is" "an instrument with a TRL value of 5") ("STI" "is" "an instrument")	(STI; is; an instrument with a TRL value of 5) (STI; is an instrument with; a TRL value)	(STI; is; instrument) (instrument; has-property; TRL value) (TRL value; has-value; 5)
STI is scheduled for acoustic testing on July 3, 2015 from 2:00PM to 6:00PM.	produced 13 triples ("STI" "is" "scheduled") ("STI" "is scheduled for" X) X = various combinations of remaining sentence	("STI" "is scheduled" "for acoustic testing on July 3 2015 from 2:00 PM") ("STI" "is scheduled" "for acoustic testing on July 3 2015 to 6:00 PM") ("STI" "is scheduled" "for acoustic testing on July 3 2015")	(STI; is scheduled; for acoustic testing) *incomplete/ missing information*: "on July 3, 2015 from 2:00PM to 6:00PM"	(STI; is; scheduled) (scheduled; for; testing) (testing; has-property; acoustic) (testing; has-value; 3 2015 July) (scheduled; from; 2:00 PM) (scheduled; to; 6:00 PM)]

Table 1: Output from Open Information Extractors. ***Red*** colored extractions are incomplete, incorrect, or noisy; ***Blue*** colored extractions need to be further granularized for ontology population.

THANK YOU!