FAKULTÄT FÜR MATHEMATIK UND INFORMATIK
Institut für Informatik
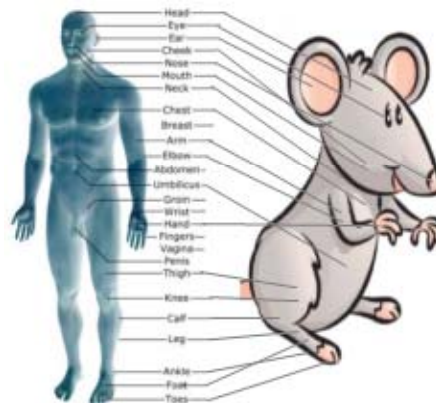Prof. Dr. Heiner Stuckenschmidt

UNIVERSITÄT
MANNHEIM

Results of the
OAEI Anatomy Track 2008

Heiner Stuckenschmidt, Christian Meilicke
KR and KM Research Group
University of Mannheim
{heiner,christian}@informatik.uni-mannheim.de

FAKULTÄT FÜR MATHEMATIK UND INFORMATIK
Institut für Informatik
Prof. Dr. Heiner Stuckenschmidt

UNIVERSITÄT
MANNHEIM

Anatomy Track

- .. is about matching ontologies describing mouse anatomy and human anatomy
- For the second time (in both 2007 and 2008) we had the same reference alignment generated by domain experts.

3304 concepts in
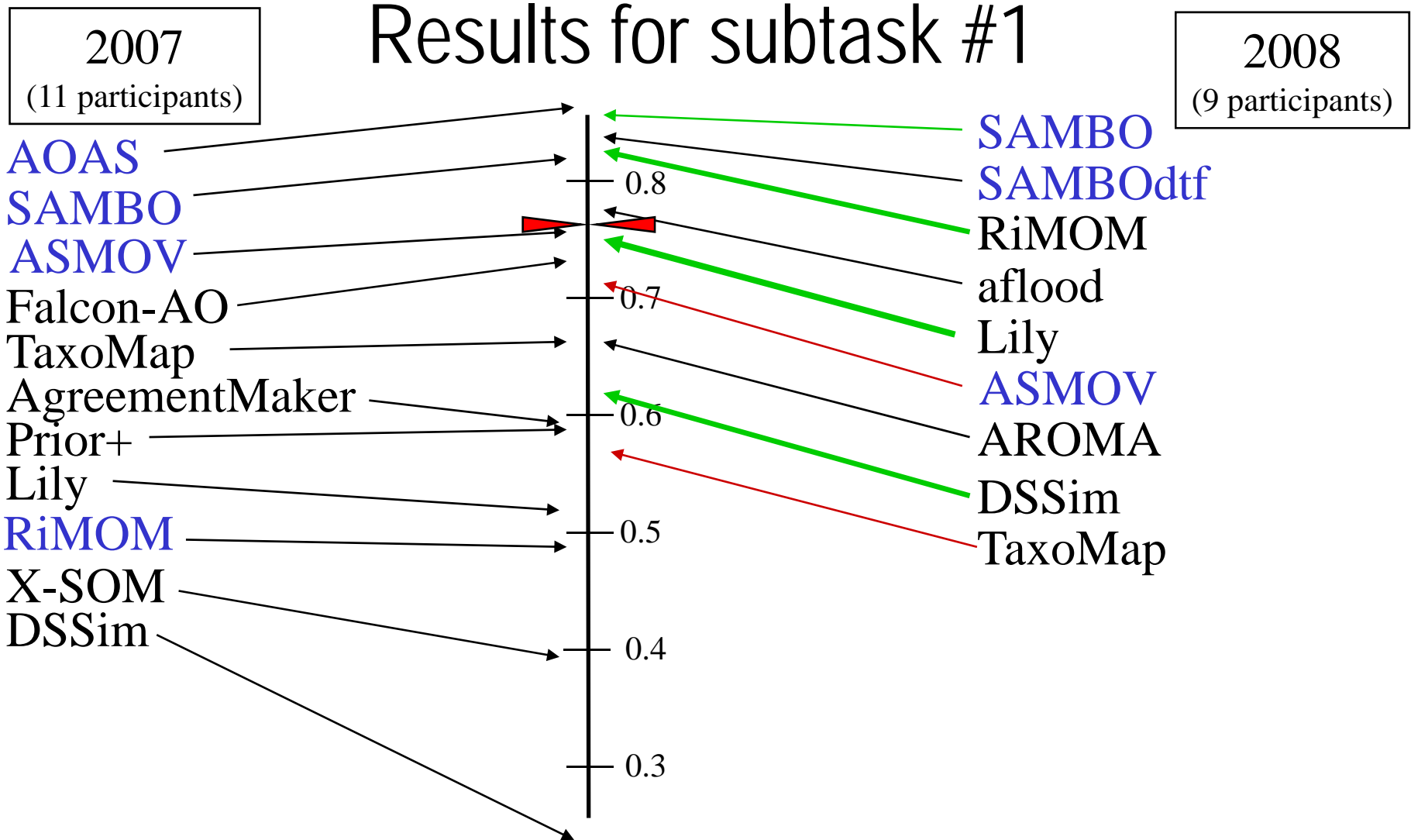human ontology

2744 concepts in
mouse ontology

1544 correspondences in reference alignment

FAKULTÄT FÜR MATHEMATIK UND INFORMATIK
Institut für Informatik
Prof. Dr. Heiner Stuckenschmidt

UNIVERSITÄT
MANNHEIM

Specifics of dataset

- Ontologies have been refactored / harmonized to in the process of creating the reference alignments
  - For example: Concepts labels have been named in similar style.

- As a result: (Unfortunately) at least ~ 60% of the reference alignment are detectable be straight forward string equality comparison after normalization (= trivial correspondences)

- Introduction of recall+ given an alignment M

- Recall+ = | (R $\cap$ M) - S | / | R - S |
  - R is the reference alignment
  - S is the alignment found by the label-equality-approach

FAKULTÄT FÜR MATHEMATIK UND INFORMATIK
Institut für Informatik
Prof. Dr. Heiner Stuckenschmidt

UNIVERSITÄT
MANNHEIM

Four Subtasks

- Task #1: Generate an alignment thats as good as possible with respect to the achieved f-value (precision an recall equally weighted)

- Task #2 and #3: Try to generate a highly precise alignment resp. an alignment with high recall

- Task #4 (new in 2008): Exploit an additionally given partial reference alignment to increase the quality of the generated alignment.

FAKULTÄT FÜR MATHEMATIK UND INFORMATIK
Institut für Informatik
Prof. Dr. Heiner Stuckenschmidt

UNIVERSITÄT
MANNHEIM

Results for subtask #1

2007
(11 participants)

2008
(9 participants)

2007 column (top to bottom):
AOAS
SAMBO
ASMOV
Falcon-AO
TaxoMap
AgreementMaker
Prior+
Lily
RiMOM
X-SOM
DSSim

2008 column (top to bottom):
SAMBO
SAMBOdtf
RiMOM
aflood
Lily
ASMOV
AROMA
DSSim
TaxoMap

Scale: 0.8, 0.7, 0.6, 0.5, 0.4, 0.3

FAKULTÄT FÜR MATHEMATIK UND INFORMATIK
Institut für Informatik
Prof. Dr. Heiner Stuckenschmidt

UNIVERSITÄT
MANNHEIM

Matcher Characteristics

	BK	Prec	Rec	Rec+	F
SAMBO	Yes	0.869	0.836	0.586	0.852
SAMBOdtf	Yes	0.831	0.833	0.579	0.832
RiMOM	No	0.929	0.735	0.350	0.821
aflood	No	0.874	0.682	0.275	0.766
Lily	No	0.796	0.693	0.470	0.741
ASMOV	Yes	0.787	0.652	0.246	0.713
...					

FAKULTÄT FÜR MATHEMATIK UND INFORMATIK
Institut für Informatik
Prof. Dr. Heiner Stuckenschmidt

UNIVERSITÄT
MANNHEIM

Recall+ for subtask #1

- No BK = detected by RiMOM, afllood, or Lily and not detected ny BK systems

- Only BK = detected by SAMBO, SAMBOdtf or ASMOV and not detected by RiMOM, aflood or Lily

Non trivial part
of the reference alignment

not found
24%

only BK
13%

no BK
13%

both
50%

FAKULTÄT FÜR MATHEMATIK UND INFORMATIK
Institut für Informatik
Prof. Dr. Heiner Stuckenschmidt
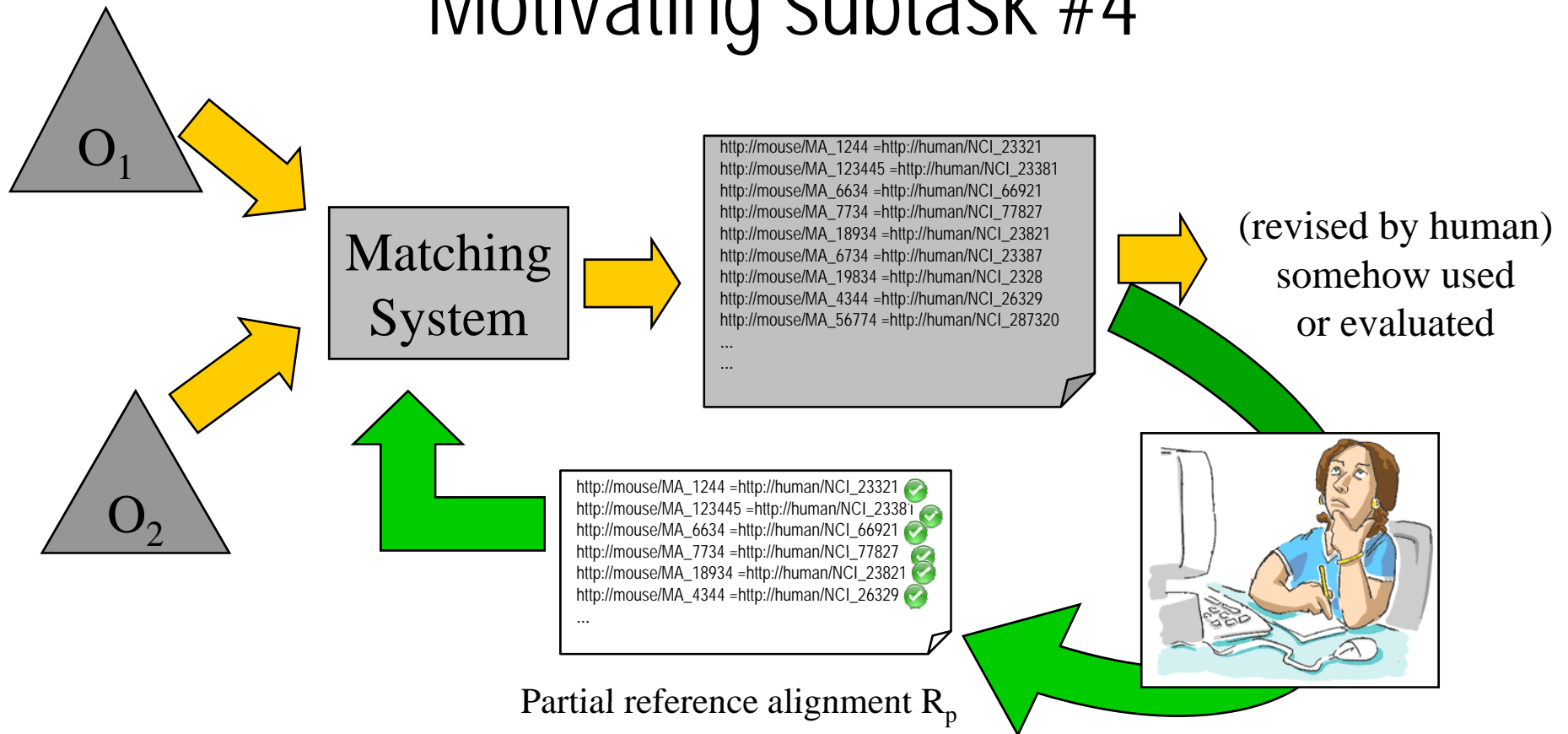
UNIVERSITÄT
MANNHEIM
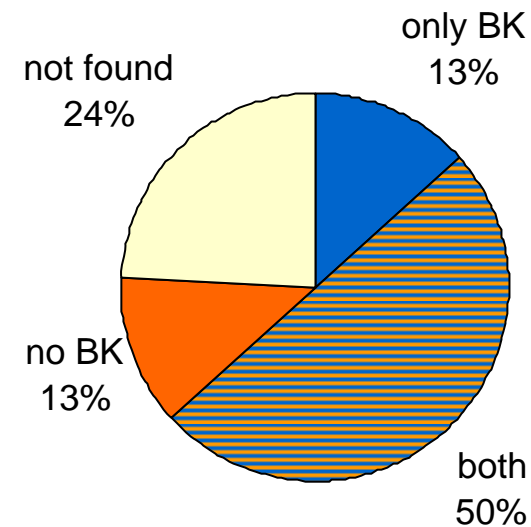
Some observations for subtask #1

- SAMBO now nearly reaches AOAS => best system in 2008

- Top system w.r.t runtime (aflood: ~ 1min compared to 4min, 20min, several hours) is also among top 4 systems!

- Most system clearly increase the quality of the generated alignments

- Parts of the overall improvement can be explained by better knowing specifics of the data set

- Some Systems not using domain specific background knowledge achieve suprisingly good results (RiMOM, aflood)

FAKULTÄT FÜR MATHEMATIK UND INFORMATIK
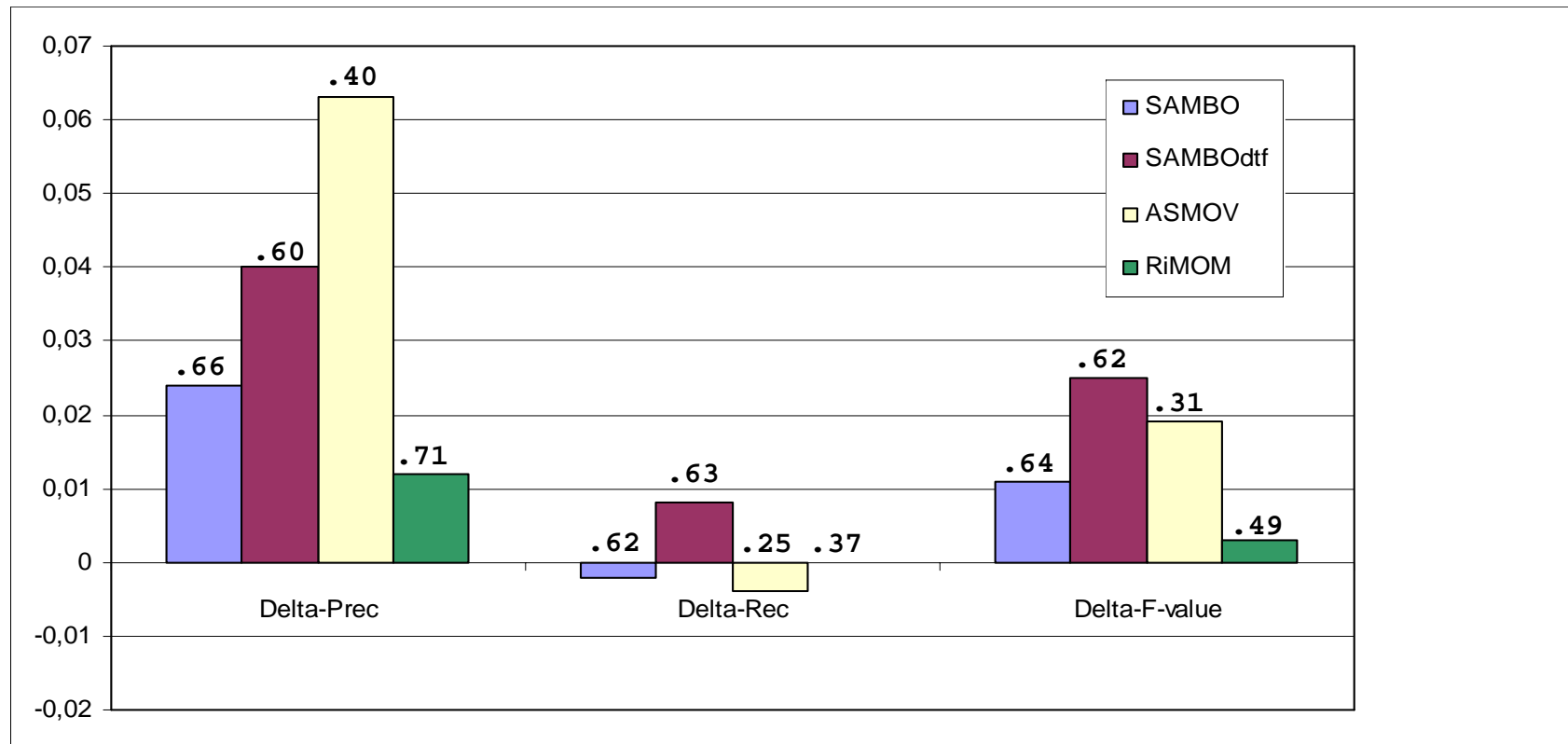Institut für Informatik
Prof. Dr. Heiner Stuckenschmidt

UNIVERSITÄT
MANNHEIM

Motivating subtask #4

O_1

O_2

Matching System

```
http://mouse/MA_1244 =http://human/NCI_23321
http://mouse/MA_123445 =http://human/NCI_23381
http://mouse/MA_6634 =http://human/NCI_66921
http://mouse/MA_7734 =http://human/NCI_77827
http://mouse/MA_18934 =http://human/NCI_23821
http://mouse/MA_6734 =http://human/NCI_23387
http://mouse/MA_19834 =http://human/NCI_2328
http://mouse/MA_4344 =http://human/NCI_26329
http://mouse/MA_56774 =http://human/NCI_287320
...
...
```

(revised by human)
somehow used
or evaluated

```
http://mouse/MA_1244 =http://human/NCI_23321 ✓
http://mouse/MA_123445 =http://human/NCI_23381 ✓
http://mouse/MA_6634 =http://human/NCI_66921 ✓
http://mouse/MA_7734 =http://human/NCI_77827 ✓
http://mouse/MA_18934 =http://human/NCI_23821 ✓
http://mouse/MA_4344 =http://human/NCI_26329 ✓
...
```

Partial reference alignment R_p

Can matching systems exploit information encoded in the partial reference alignment?

FAKULTÄT FÜR MATHEMATIK UND INFORMATIK
Institut für Informatik
Prof. Dr. Heiner Stuckenschmidt

UNIVERSITÄT
MANNHEIM

The partial reference alignment R_p

- Consists of 988 correspondences
  - Trivial part of reference alignment (= 934 correspondences) and
  - 54 non trivial correspondences (subset of correct correspondences detected by systems in 2007)
  - $\Rightarrow$ 535 correct correspondences remain unknown, referred to as R_u
- Evaluation results on following slide are based on comparing $M - R_p$ with R_u given resp. not given R_p

FAKULTÄT FÜR MATHEMATIK UND INFORMATIK
Institut für Informatik
Prof. Dr. Heiner Stuckenschmidt

UNIVERSITÄT
MANNHEIM

Comparing the results of #1 and #4

FAKULTÄT FÜR MATHEMATIK UND INFORMATIK
Institut für Informatik
Prof. Dr. Heiner Stuckenschmidt

UNIVERSITÄT
MANNHEIM

Conclusions for subtask #4

- 4 of 9 matching system have participated, all with positive results ☺

- Overall we measured increased precision with only minor changes in recall
  - SAMBOdtf managed to increase both precision and recall

- Further consideration have to focus on specific strategies used and their effects
  - Look into the "matcher-black-box"

FAKULTÄT FÜR MATHEMATIK UND INFORMATIK
Institut für Informatik
Prof. Dr. Heiner Stuckenschmidt

UNIVERSITÄT
MANNHEIM

New matching tasks?

- We would like to extend the anatomy track with additional matching tasks!
  - Any new ideas?
  - Biomedical ontologies + reference alignments used in your own experiments?

- => Contact us!

FAKULTÄT FÜR MATHEMATIK UND INFORMATIK
Institut für Informatik
Prof. Dr. Heiner Stuckenschmidt

UNIVERSITÄT
MANNHEIM

Thanks for your attention

Any questions?