Journée Commune AFIA - ARIA| Paris, December 2017

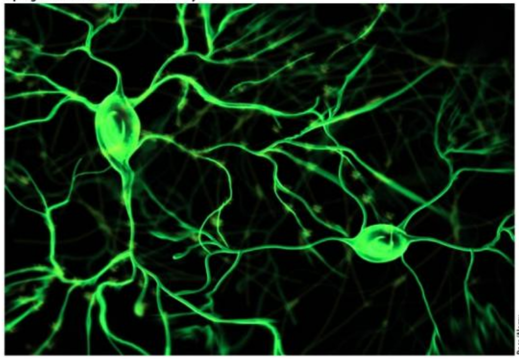
Learning with Knowledge Graphs: From Medical Decision Support to Human Perception and Memory

Volker Tresp

Siemens AG and Ludwig Maximilian University of Munich

Knowledge Graphs and Learning with Knowledge Graphs

Optogenetic control of neural activity



Deco/Alamy

Why Machine Learning Should Care about Knowledge Graphs

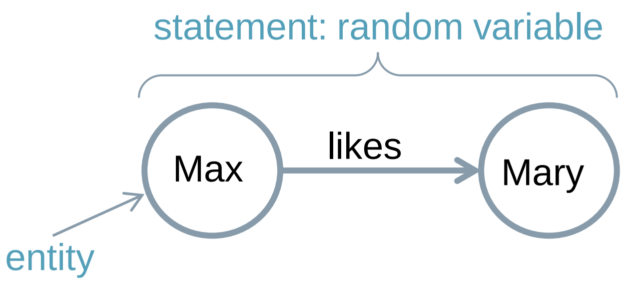
- Machine Learning is great in stimulus-response type of learning (classification, prediction, …)
- How can machine learning exploit the fact that “Obama is ex president of the United States?”
- The field of *Knowledge Representation (KR)* builds on Relational Databases and Knowledge Graphs

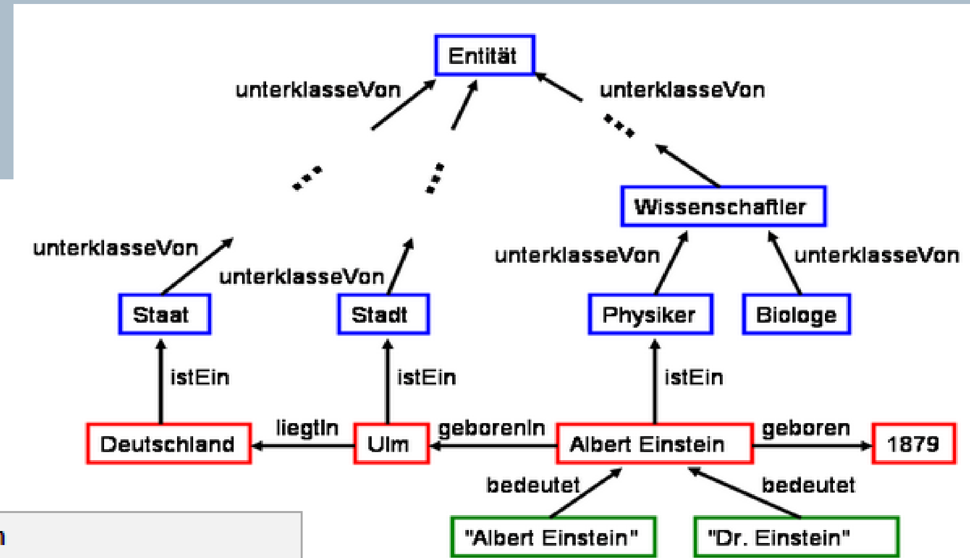
„You must be ex president Obama!“

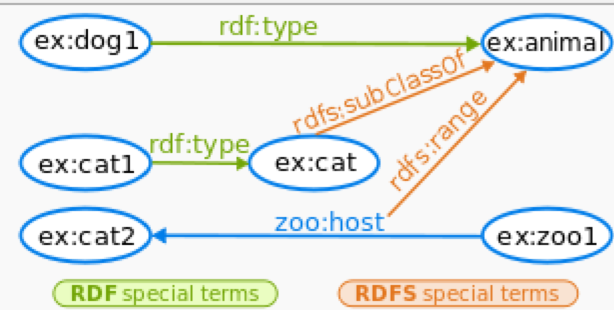
„How is your wife Michelle?“

(induction: “likely to be male and white”)

Triple Graphs





In english	The graph
• Dog1 is an animal • Cat1 is a cat • Cats are animals • Zoos host animals • Zoo1 hosts the Cat2	

RDF/turtle

```
@PREFIX rdf:   <http://www.w3.org/1999/02/22-rdf-syntax-ns#> .
@PREFIX rdfs:   <http://www.w3.org/2000/01/rdf-schema#> .
@PREFIX ex:   <http://example.org/> .
@PREFIX zoo:   <http://example.org/zoo/> .
ex:dog1     rdf:type          ex:animal .
ex:cat1     rdf:type          ex:cat .
ex:cat      rdfs:subClassOf   ex:animal .
zoo:host    rdfs:range        ex:animal .
ex:zoo1     zoo:host          ex:cat2 .
```

- The formal language used in KR is extremely difficult to understand for outsiders
- Fortunately a triple-oriented Knowledge Graph is very simple to understand and is sufficient for many (scalable) applications
- A Knowledge Graph as a link between cultures: KR and Machine Learning
- Brings order into things with surprisingly simple constructs

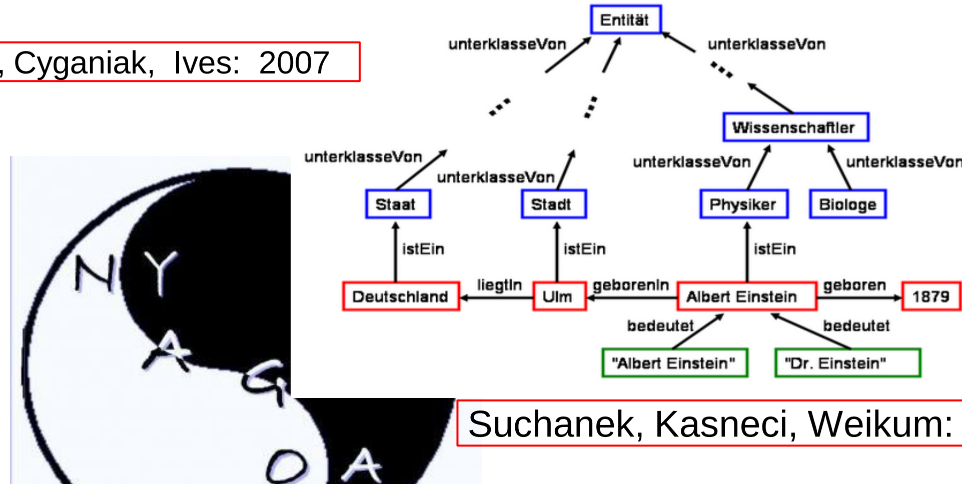
For the First Time there Exist Sizable Knowledge Graphs: DBpedia, YAGO, Freebase, Nell, Google Knowledge Graph, Microsoft's Satori, Facebook's Entity Graph

Auer, Bizer, Kobilarov, Lehmann, Cyganiak, Ives: 2007

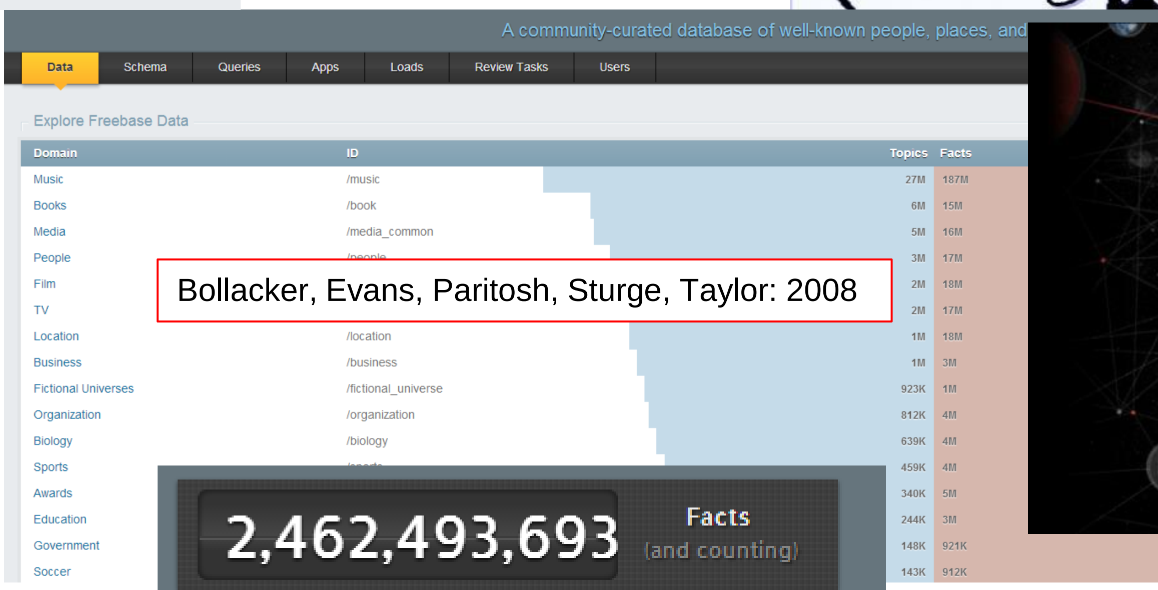
About / News

Datasets

DBpedia is a community effort to extract structured information from Wikipedia and to make this information available on the Web. DBpedia allows you to ask sophisticated queries against Wikipedia, and to link other data sets on the Web to Wikipedia data.

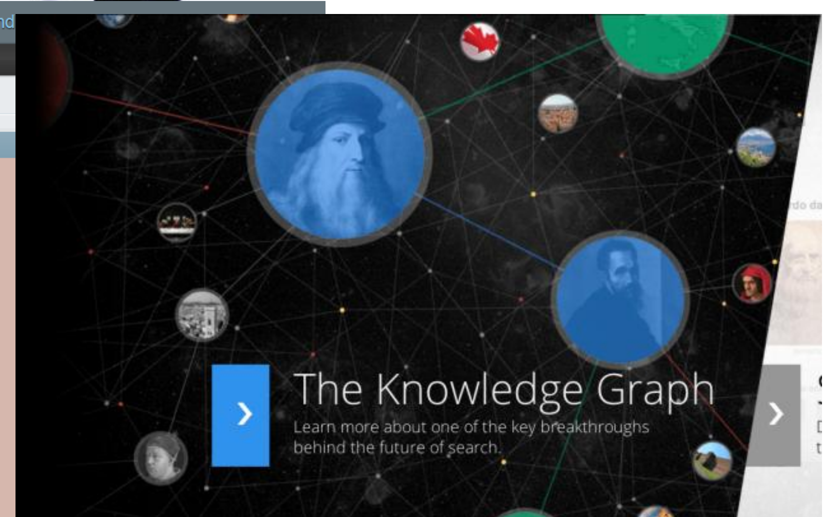


Suchanek, Kasneci, Weikum: 2007



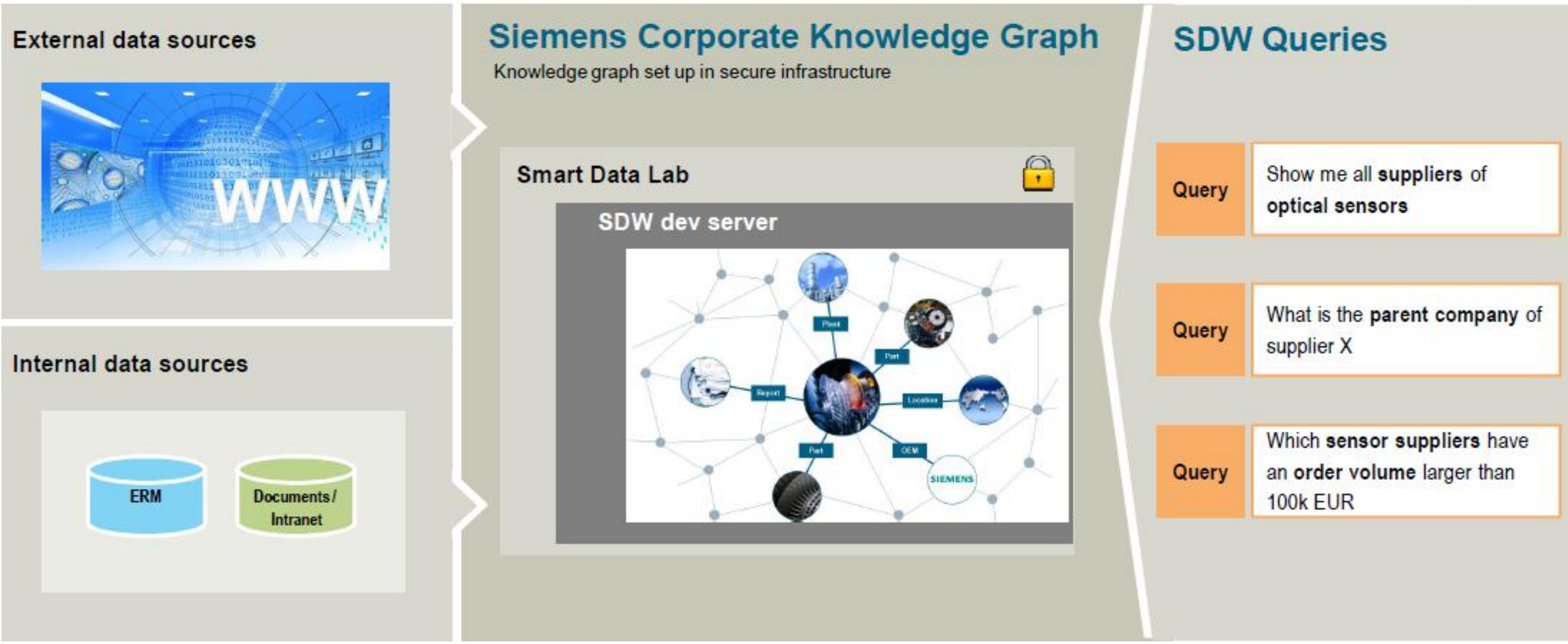
Domain	ID	Topics	Facts
Music	/music	27M	187M
Books	/book	6M	15M
Media	/media_common	5M	16M
People		3M	17M
Film		2M	18M
TV		2M	17M
Location	/location	1M	18M
Business	/business	1M	3M
Fictional Universes	/fictional_universe	923K	1M
Organization	/organization	812K	4M
Biology	/biology	639K	4M
Sports		459K	4M
Awards		340K	5M
Education		244K	3M
Government		148K	921K
Soccer		143K	912K

Bollacker, Evans, Paritosh, Sturge, Taylor: 2008



>100B Facts

Not just Google ….

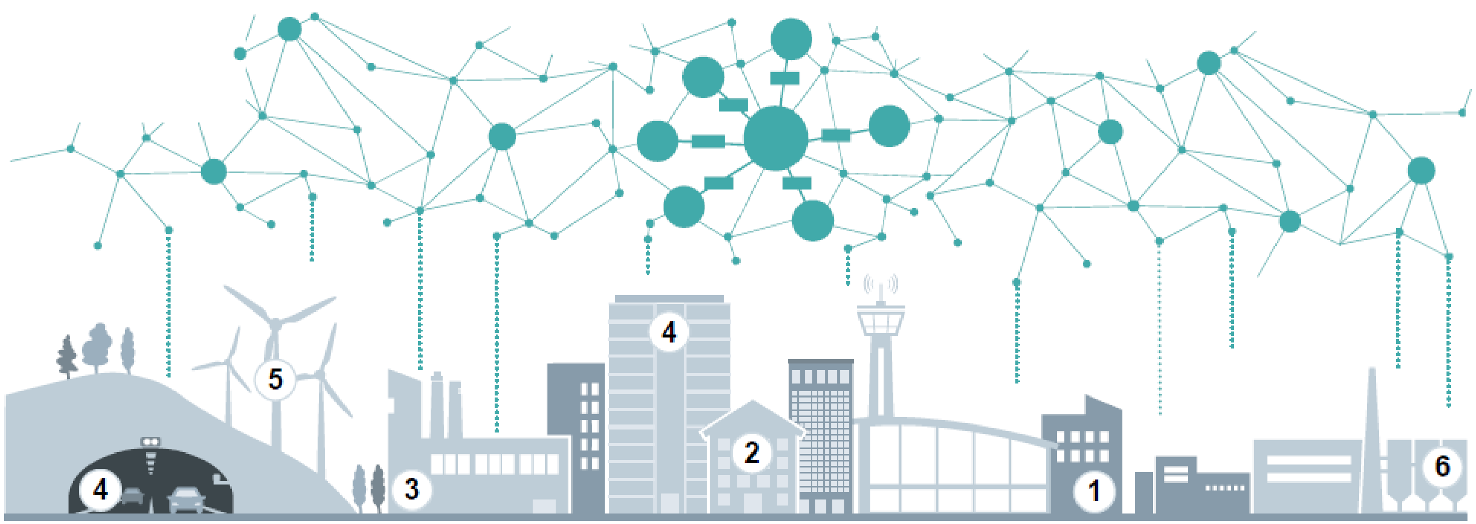


- Data and knowledge integration
- Question-answering
- Support for search
- Statistics, visualization, trend analysis

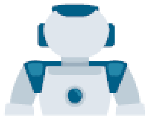
One day in the life of an artificial assistant @ Siemens

SIEMENS
Ingenuity for life

Industrial Knowledge Graph

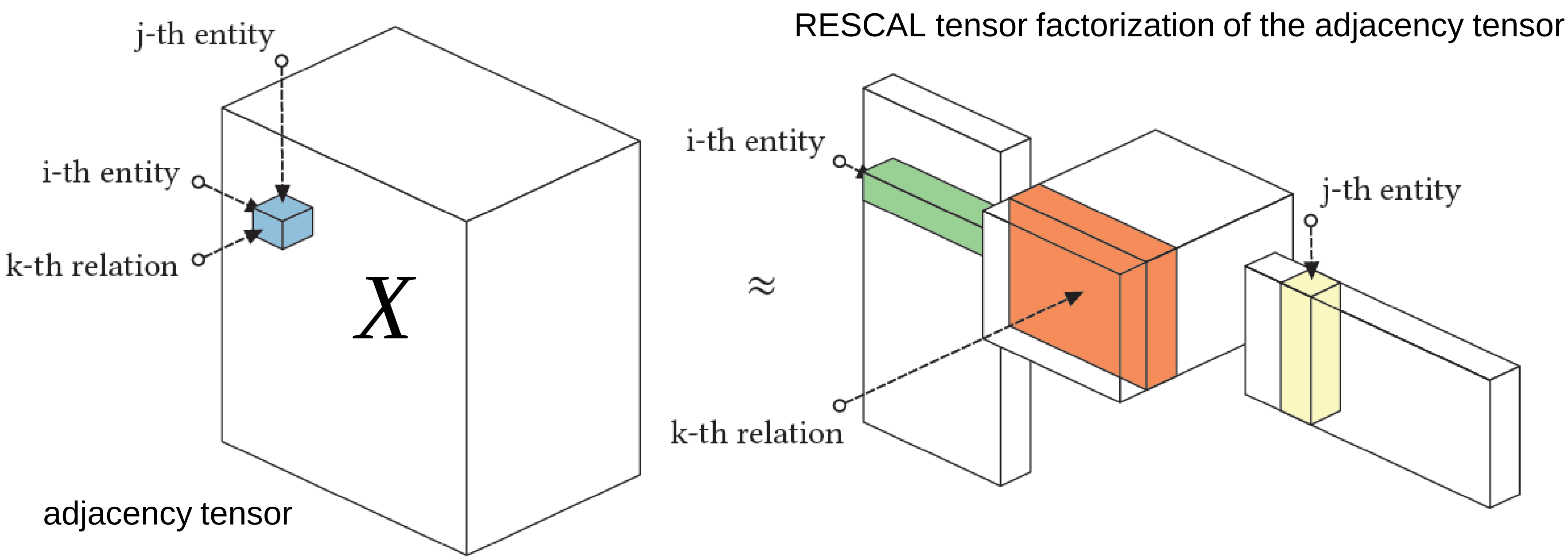


1. **09:00 – Understand**
   Tender Search
2. **11:00 – Offer**
   Configure Solution
3. **12:00 – Produce**
   Flexible Manufacturing
4. **13:00 – Supply**
   Supply Chain Mgmt.
5. **15:00 – Maintain**
   Maintenance planning
6. **18:00 - Repair**
   Advanced Diagnostics



AI Algorithms

Machine Learning:
Generalization via Tensor Decomposition



Training Data:

$$x_{s,p,o} = 1$$ If *(s,p,o)* is known to be true

$$x_{s,p,o} = 0$$ otherwise

After factorization (RESCAL: constr. Tucker2):

$$P((s,p,o)) = \mathrm{sig}(\theta_{s,p,o})$$

$$\theta_{s,p,o} = \sum_{r_1} \sum_{r_3} a_{e_s,r_1} a_{e_o,r_3} g(r_1, p, r_3)$$

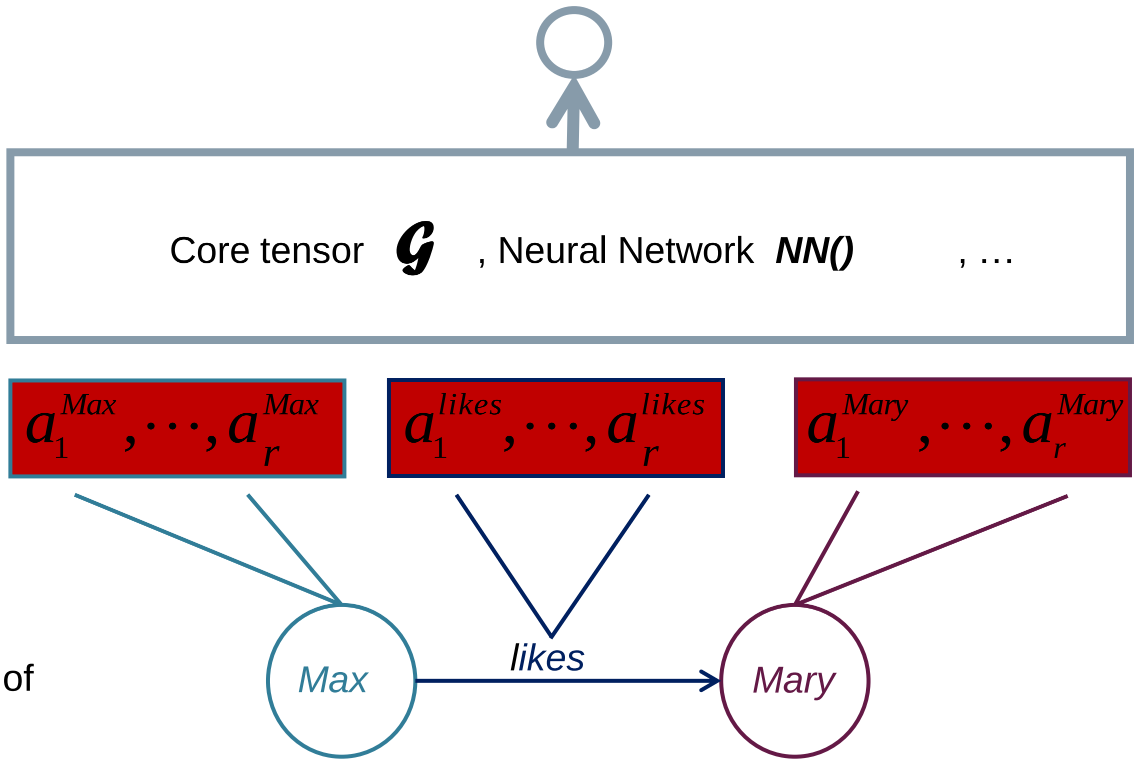
$$\Theta = \mathcal{G} \times_1 A \times_2 A$$

Nickel, Tresp, Kriegel. A Three-Way Model for Collective Learning on Multi-Relational Data. ICML 2011

Tensor Factorization as Representation Learning

- We maintain that an adjacency tensor is the appropriate representation
- Different forms of representation learning

$$P(\quad (Max, likes, Mary) = true \quad)$$

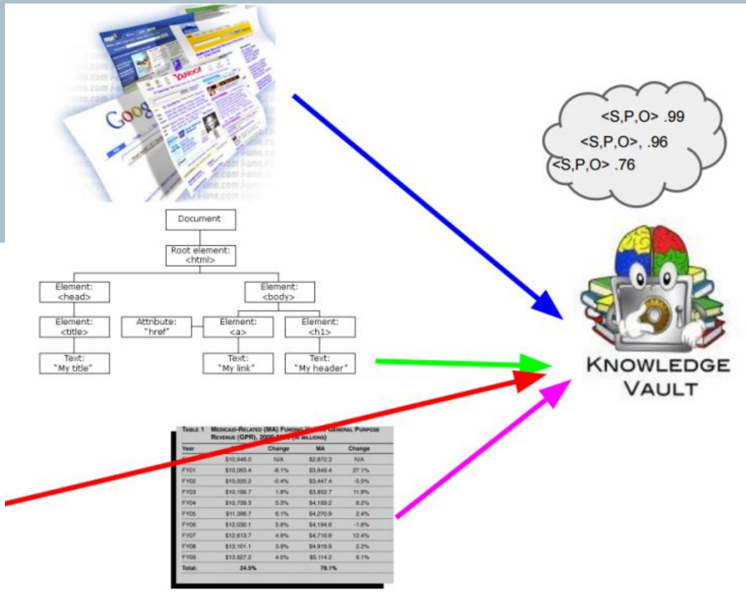


<u>Representation Learning:</u>

Shared latent representations of *Max, likes, and Mary*

We are able to predict all typed links (> 100 types) between several million of nodes!

The Google Knowledge Vault



- **Extractors**
  - Fused Extractors: AUC = 0.927
- **Graph-based priors**
  - Path ranking algorithm (PRA) (AUC = 0.884)
  - Factorization approach (multiway Neural Network) (AUC = 0.882)
  - Fused Prior: AUC 0.911
- **Knowledge fusion**
  - This system computes the probability of a triple being true, based on agreement between different extractors and priors
  - Combined Overall System: AUC = 0.947

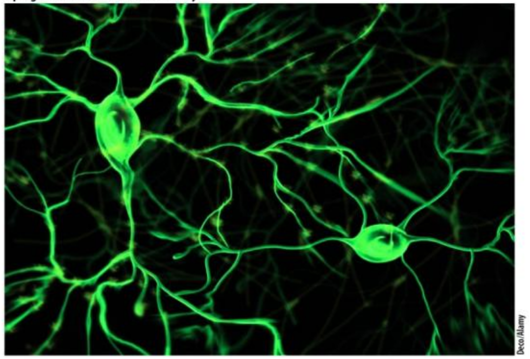
The number of high-confidence facts increases from **100M to about 271M**

Dong, Gabrilovich, Heitz, Horn, Lao, Murphy, Strohmann, Sun, Zhang. Knowledge Vault: A Web-scale Approach to Probabilistic Knowledge Fusion. KDD, 2014

Nickel, Murphy, Tresp, Gabrilovich. A Review of Relational Machine Learning for Knowledge Graphs: From Multi-Relational Link Prediction to Automated Knowledge Graph Construction. Proceedings of the IEEE, (invited paper), 2016.

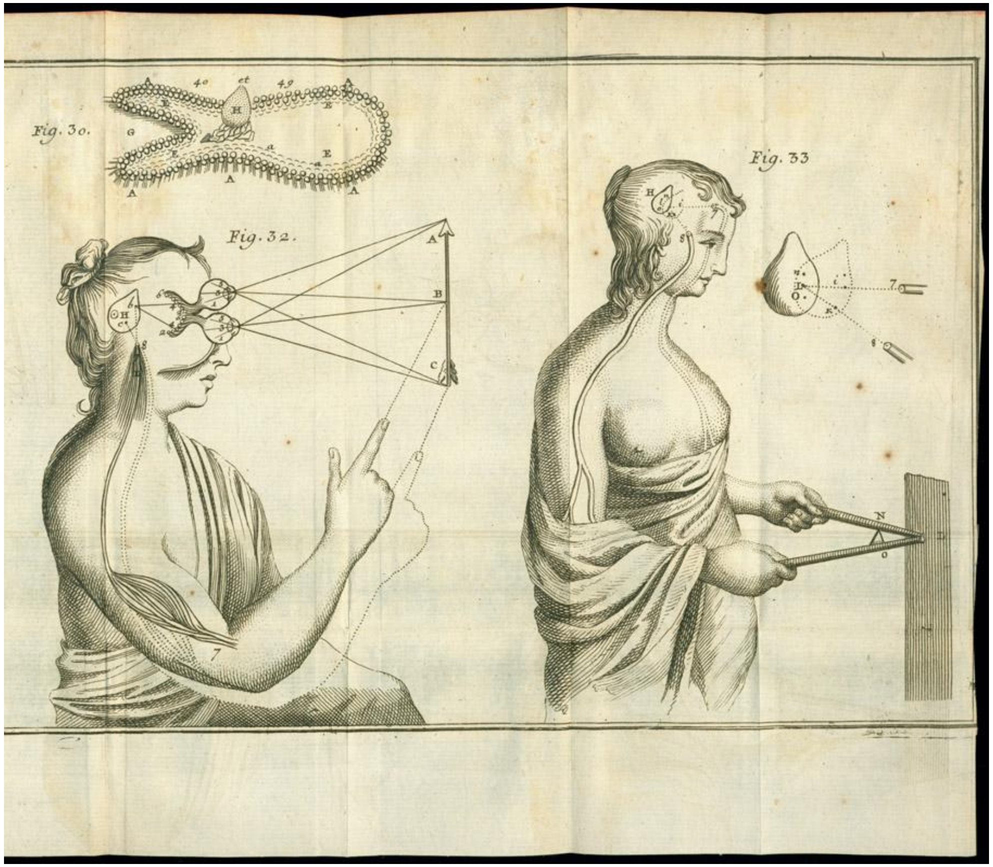
A role for Deep Learning

Optogenetic control of neural activity

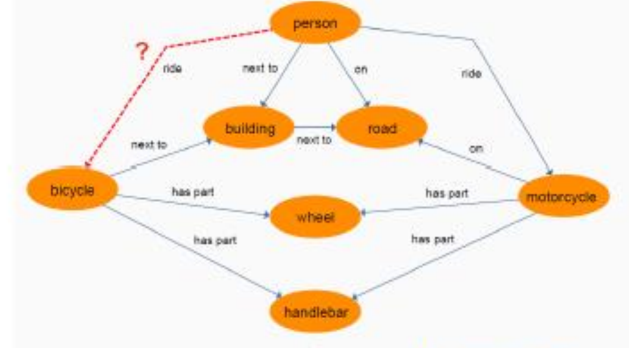


Smart Perception: Integrating Knowledge Graphs with Deep Learning

- Understanding the world means knowing the world



Smart Perception:
Integrating Knowledge Graphs with Deep Learning



Learning Knowledge Graph

Deep Learning **Deep Learning**

S = Person
P = nextTo
O= Surfboard

- By using a KG prior, we obtained better results than the Stanford group: *Lu, Krishna, Bernstein, Fei-Fei, 2016*
- Best student paper at the ISWC 2017
- *MLwin - Maschinelles Lernen mit Wissensgraphen, Tresp, Schütze, Weikum, Cremers, et al.*

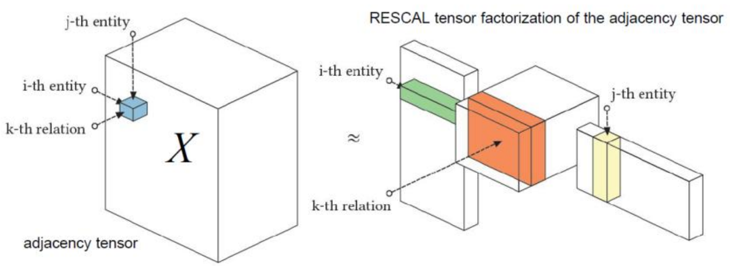
Unrestricted © Siemens AG 2017

Baier, Ma, Tresp. Improving Visual Relationship Detection using Semantic Modeling of Scene Descriptions, ISWC 2017

Table 1: Results for visual relationship detection. We report Recall at 50 and 100 for four different validation settings.

Task	Phrase Det.		Rel. Det.		Predicate Det.		Triple Det.	
Evaluation	R@100	R@50	R@100	R@50	R@100	R@50	R@100	R@50
Lu et al. V [14]	2.61	2.24	1.85	1.58	7.11	7.11	2.68	2.30
Lu et al. full [14]	17.03	16.17	14.70	13.86	47.87	47.87	18.11	17.11
RESCAL	19.17	18.16	16.88	15.88	52.71	52.71	20.23	19.13
MultiwayNN	18.88	17.75	16.65	15.57	51.82	51.82	19.76	18.53
ComplEx	19.36	18.25	17.12	16.03	53.14	53.14	20.23	19.06
DistMult	15.42	14.27	13.64	12.54	42.18	42.18	16.14	14.94

Learning Knowledge Graphs



Training Data:

$x_{s,p,o} = 1$ If (s,p,o) is known to be true

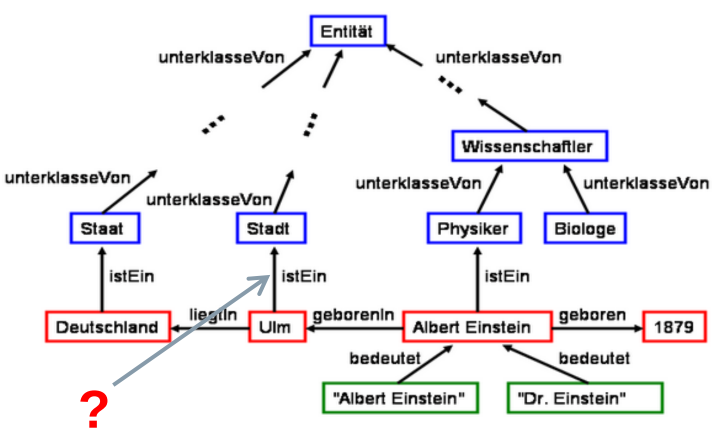
$x_{s,p,o} = 0$ otherwise

After factorization (RESCAL2: constr. Tucker2):

$$P((s,p,o)) = \mathrm{sig}(\theta_{s,p,o})$$

$$\theta_{s,p,o} = \sum_{r_1} \sum_{r_3} a_{e_s,r_1}\, a_{e_o,r_3}\, g(r_1,p,r_3)$$

$$\Theta = G \times_1 A \times_2 A$$



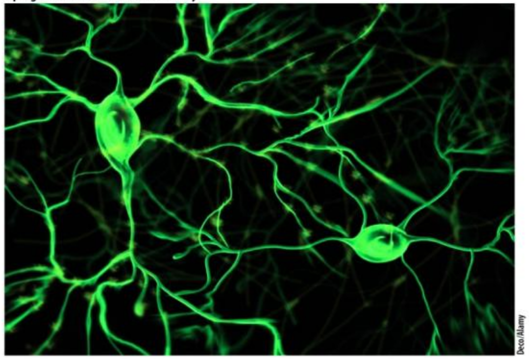
- Inferential queries
  - What disease does Jack likely have?
- Automatic filling of KG
  - Knowledge Vault projects
- Learning Database
- Detection of KG errors
- KG priors to understand text and images
- Use as background information (compressed as latent factors) that can be used in other applications (predictions, decision support)

- RESCAL works! But no one tool fits all
- Factorization is good at patterns with one universal quantifier (but potentially many existential quantifiers)

*Nickel, Jiang, Tresp. Reducing the Rank of Relational Factorization Models (NIPS*2014), 2014*

Semantic Memory (sKG) and Episodic Memory (eKG)

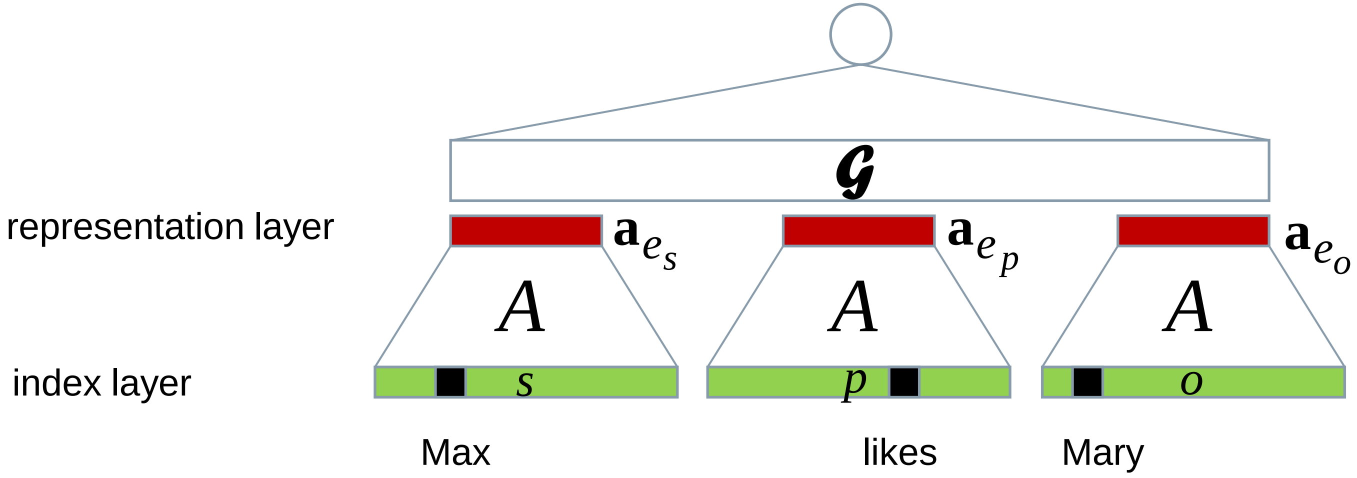
Optogenetic control of neural activity



Semantic Memory: *Facts We Know*

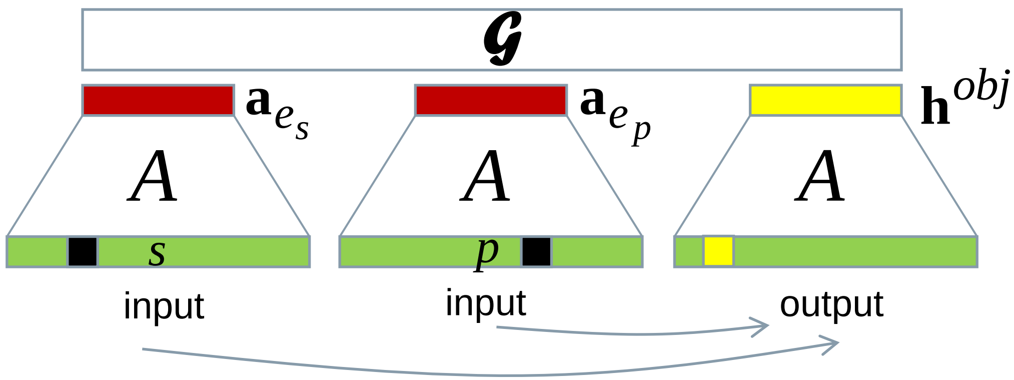
- A KG is a technical realization of a semantic memory
- A learned KG model is an efficient store that performs inductive inference

$$\theta_{s,p,o} = f(\mathbf{a}_{e_s}, \mathbf{a}_{e_p}, \mathbf{a}_{e_o})$$



Mathematically Equivalent “Conditional” Representation Biologically more Plausible

- Nice feature of tensor models:
  - Mathematically equivalent view (translational view) for tensor models: mapping of the latent representations of subject and predicate should be close to the latent representation of the object

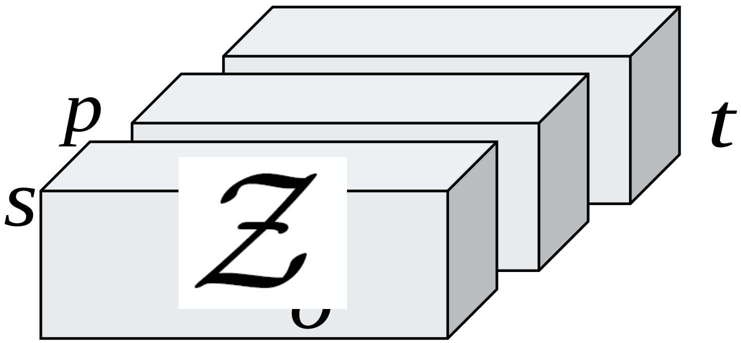


$$\theta_{s,p,o} = (\mathcal{G} \times_1 \mathbf{a}_s \times_2 \mathbf{a}_p) \cdot \mathbf{a}_o = \mathbf{a}_o^T \mathbf{h}_{s,p}^{obj}$$

Episodic Memory: *Facts We Remember*

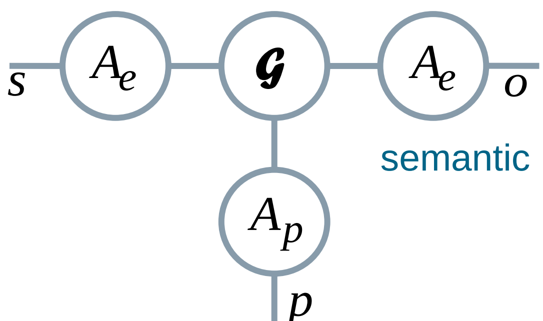
- Many applications (e.g., healthcare) require a notion of time
- An episodic KG is a technical realization of an episodic memory

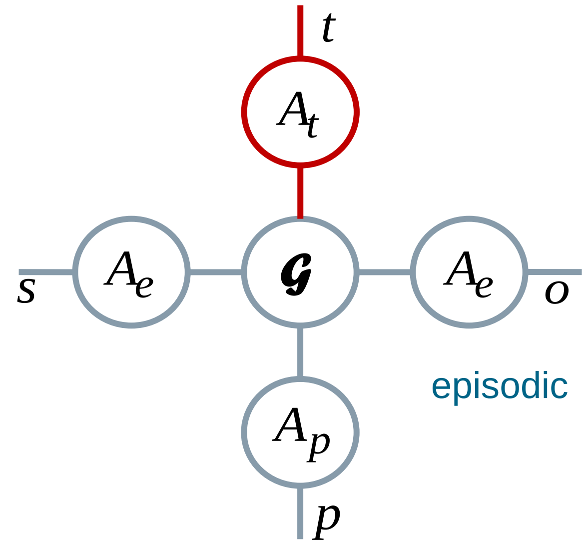
4-way episodic event tensor



Stands for anything that happened at time *t*

$$\theta^{episodic}_{s,p,o,t} = f^{episodic}(\mathbf{a}_{e_s}, \mathbf{a}_{e_p}, \mathbf{a}_{e_o}, \mathbf{a}_{e_t})$$

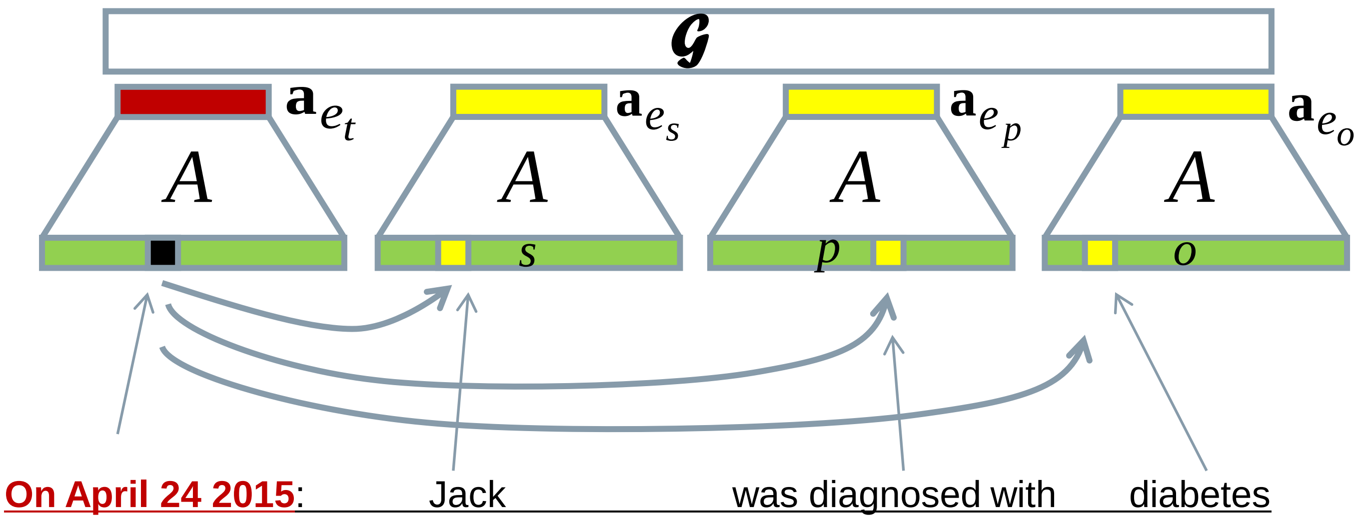




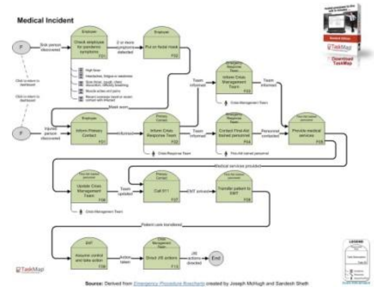
Tensor Networks (Tucker Models)

Learned Episodic Memory

Episodic memory



Semantic and Episodic Memories for Decision Modeling



- *Esteban, Staeck, Yang, Tresp. IEEE ICHI, 2016*
- *Yang, Fasching, Tresp. MUCMD 2017*
- *Yang, Fasching, Tresp. IEEE ICHI 2017*

Semantic Knowledge Graph

- What do I know about the patient?

Episodic Knowledge Graph

- What happened to the patient?

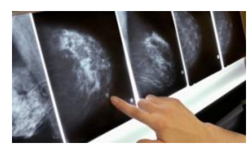
- Decision Modeling
- Endpoint Prediction
- sKG, eKG, LSTM, ..

SIEMENS
Ingenuity for life

Decision Support

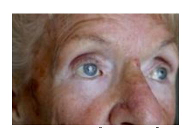
- Best treatment?

Smart Data



Universitätsklinikum Erlangen

CHARITÉ
UNIVERSITÄTSMEDIZIN BERLIN



LMU KLINIKUM DER UNIVERSITÄT MÜNCHEN

age-related macular degeneration

Smart Data in Healthcare

Detailed Information
about each individual patient
(more dimensions; over time)

Precision medicine
- Information overload!
- Need for IT support and automation

Information
about many patients
(more instances)

Learning healthcare system
- Descriptive A. (what has happened?)
- Diagnostic A. (why? Insight!)
- Predictive A. (what will happen?)
- Prescriptive A. (what should be done?)

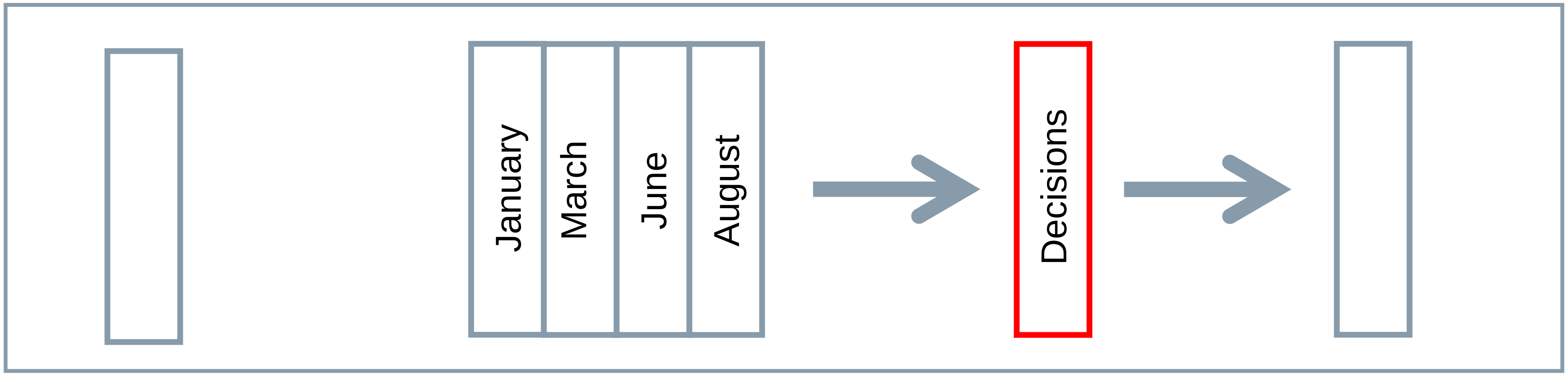
Machine Learning Approach: Predictive Modeling

Principle: Decision Modelling

- „The knowledge of the physicians ---including years of training, experience, publications they read--- is only relevant in as much as it influences medical decisions"
- And it is reflected in their decisions!

Principle: Endpoint Prediction

- Endpoint prediction can serve many purposes
- Decision support: „Propose decisions which are optimal under the predictive model to reach best end points"



Background (sKG)

- Age, gender, preconditions, …, primary tumor, history of metastasis before the study

Sequential (eKG)

- measurements, decisions

Recurrent Deep Learning

Endpoint

- "progression-free survival"

Data and Static Features

Static features	Feature names and dimensions	
1) Basic	Age	1
	Height	1
	HRT (Hormone Replacement Therapy)	5
	parity	9
	Mother BC	3
	Sister BC	6
	Menstruation	1
2) Primary Tumor	Type	3
	Total eval. of the malignancy	8
	Total eval. of axilla	4
	TAST eval. of the malignancy	8
	TAST eval. of axilla	4
	Mammography eval. of the malignancy	8
	Mammography eval. of axilla	4
	Ultrasound eval. of the malignancy	8
	Ultrasound eval. of axilla	4
	MRI eval. of the malignancy	8
	MRI eval. of axilla	8
	Metastasis staging	4
	Ever neoadjuvant therapy	4
	Ever surgery	4
3) History of metastasis	Lungs	1
	Liver	1
	Bones	1
	Brain	1
	Other	10
Total	26	118

Patient features:

Static (in total 118 features)

- (1) Basic information
- (2) Primary
- (3) History of metastasis before the study

Sequential Features

Patient features:

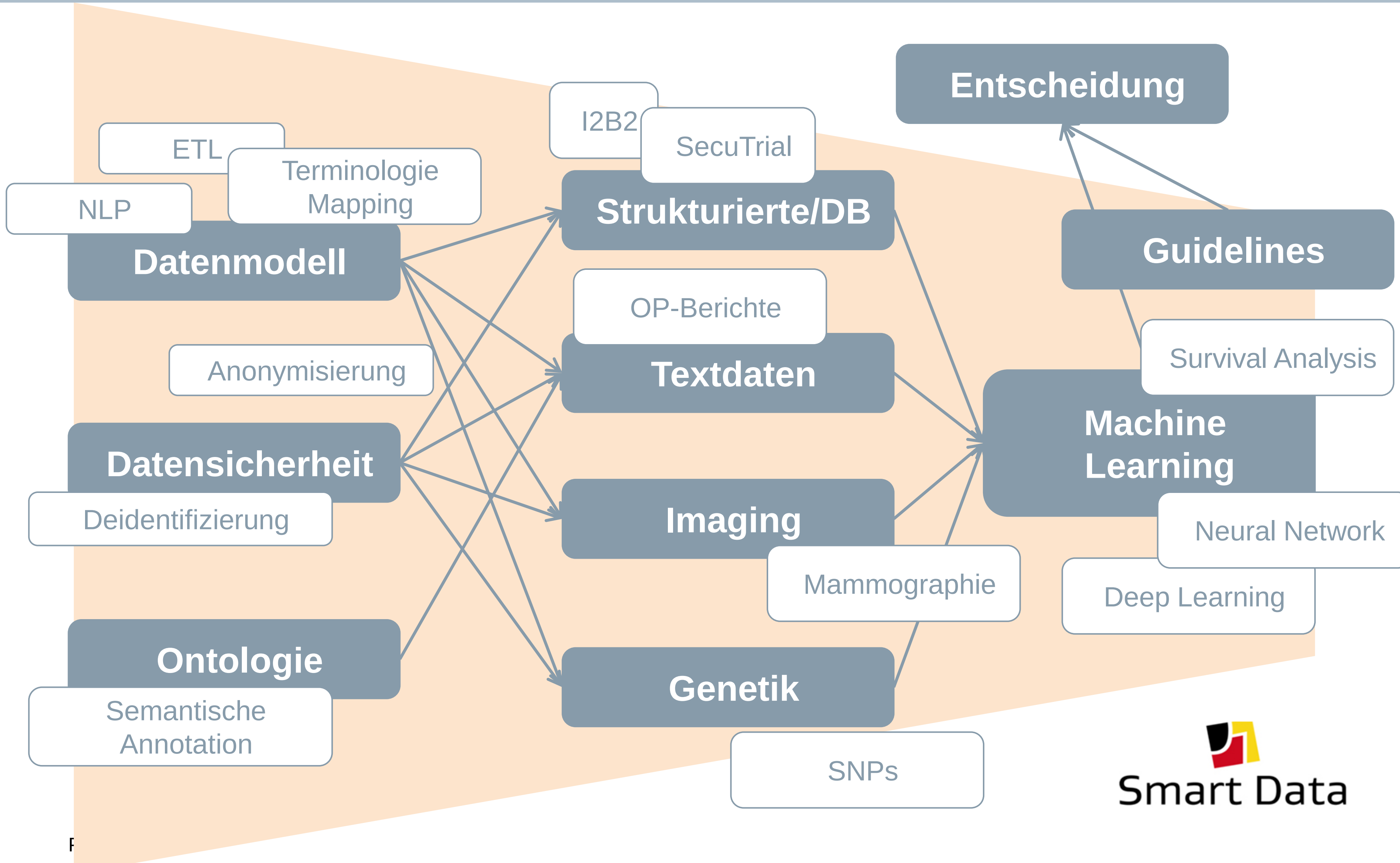
Sequential (in total 189 features)

- (4) local recurrence
- (5) metastasis
- (6) clinical visits
- (7) radiotherapy
- (8) systemic therapy
- (9) surgery

Challenge: The length of the sequence in each patient case varying from 0 to 35

Sequential features	Feature names and dimensions	
4) Local Recurrences	Location	4
	Type	3
5) Metastasis Evaluation	Total	6
	Lungs	9
	Liver	9
	Bones	9
	Brain	9
	Lymph	9
	Skin	9
	Ovum	9
	Soft tissue	8
	Kidney	8
	Pleural cavity	8
	Thorax	8
	Muscle	8
	Periosteum	8
	Other	8
6) Visits	Therapy situation	12
	ECOG Life status	6
7) Radiation	Type	3
	Intention	3
8) Systemic	Type	6
	Intention	13
	Ref. to an surgery	4
	Reason of termination	6
9) Surgery	Type	10
Total	26	189

Integrated Decision Support



Smart Data

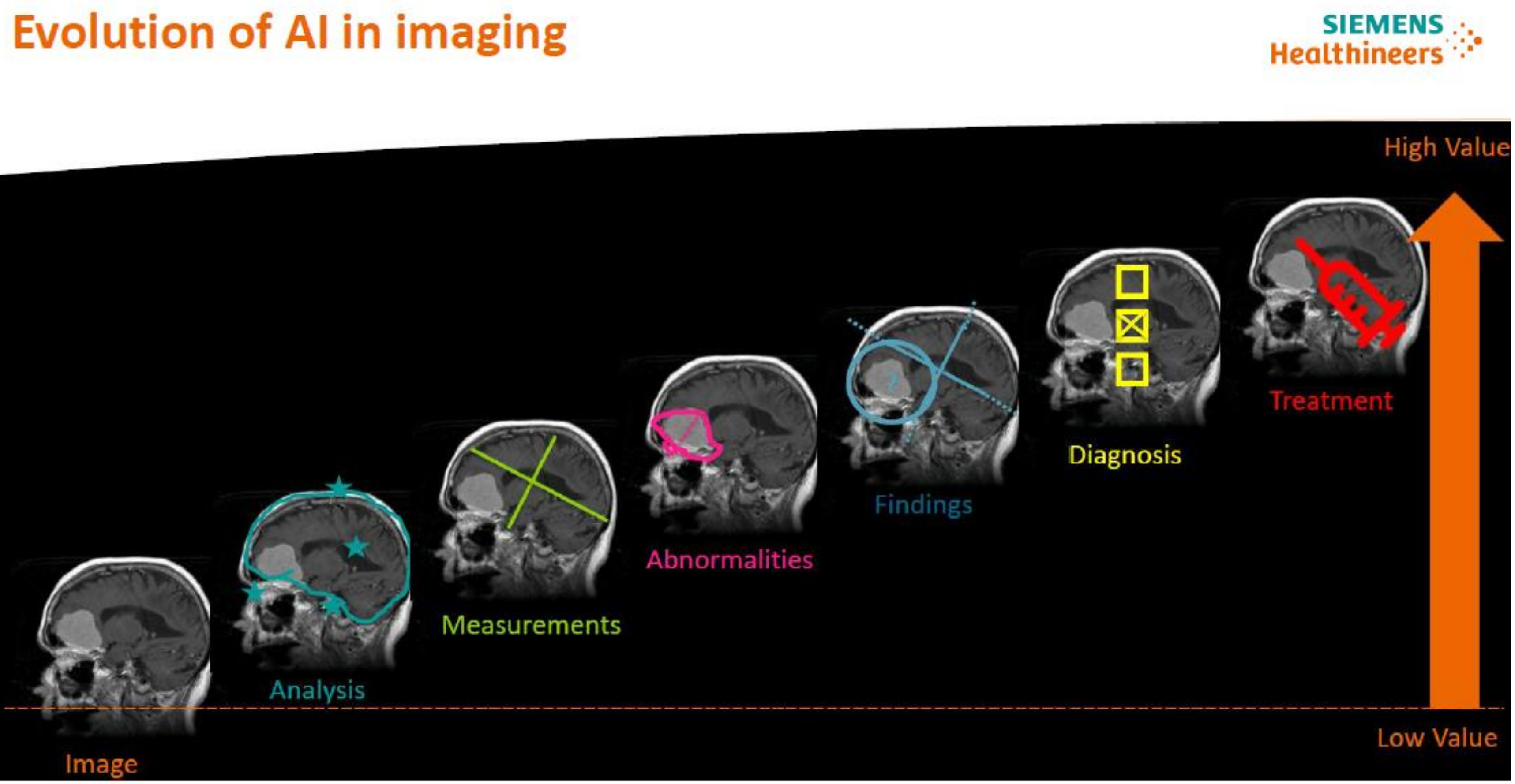
Clinical Feasibility and Accuracy

- The improvements are statistically significant, but what does it mean?

	Acceptable alternative	Don‘t agree	Don‘t agree at all
Re-Tumorboard	11%	64%	23%
Machine Decision	33%	58%	8%

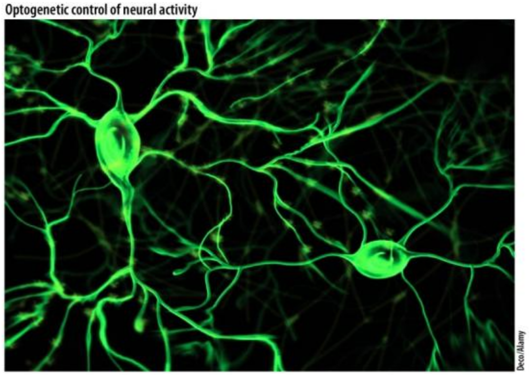
AI in Radiology

Increasing relevance for Siemens Healthcare



Perception and Relationships to Human Memory

Optogenetic control of neural activity



The Amazing Brain

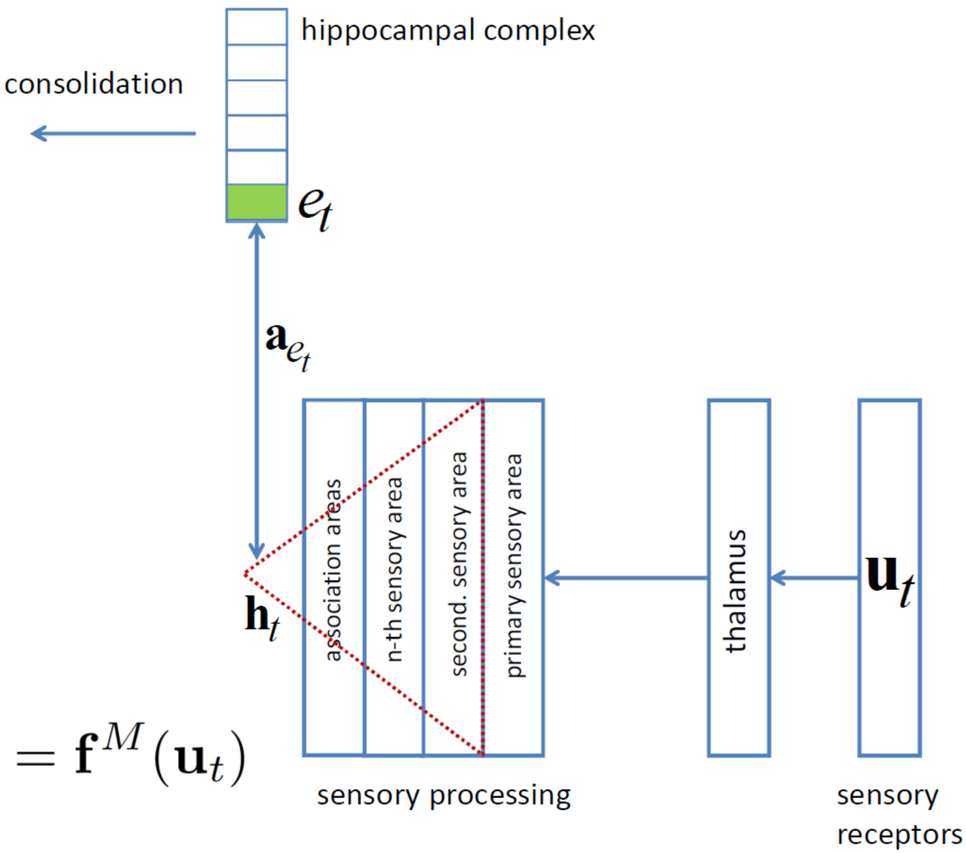
Sensor Processing
- Fast, skillful reaction
- Human declarative capabilities
  - Deep understanding of sensory inputs; **declarative decoding**; with a link to language

Episodic memory ("facts we remember")
- Recall a sensory impression of past events
- Human **declarative memory**

Semantic memory ("facts we know")
- "Obama is ex-president of the United States"; "Munich is in Bavaria"
- Human **declarative memory**

More: decisions; prediction; reasoning; action; learning from episodes, ….

Hippocampal Memory Indexing Theory: From Sensory Input to Epsiodic Memory

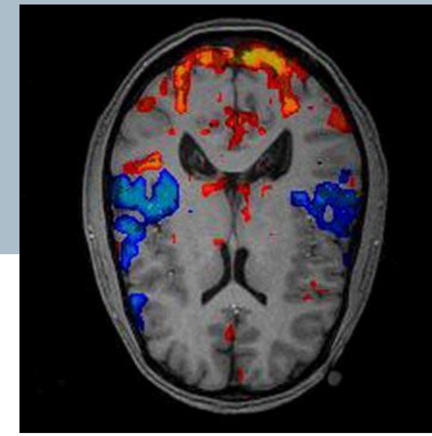


$$\mathbf{h}_t = \mathbf{f}^M(\mathbf{u}_t)$$

Teyler & DiScenna, 1986; Teyler & Rudy, 2007

- The latent representation for time is represented in the higher order layers of sensory processing / association cortex
- For meaningful episodes, an index for the time instance is generated in the hippocampal area
- Engram: index & representation $(e_t, \mathbf{a}_{e_t})$
- Recollection (internal stage, subsymbolic) by back projection: $(e_t, \mathbf{a}_{e_t})$ is reactivated, including the bound neocortical traces
- *But: not a model for explicit memory!*

Starting Point: Sensor Hierarchy

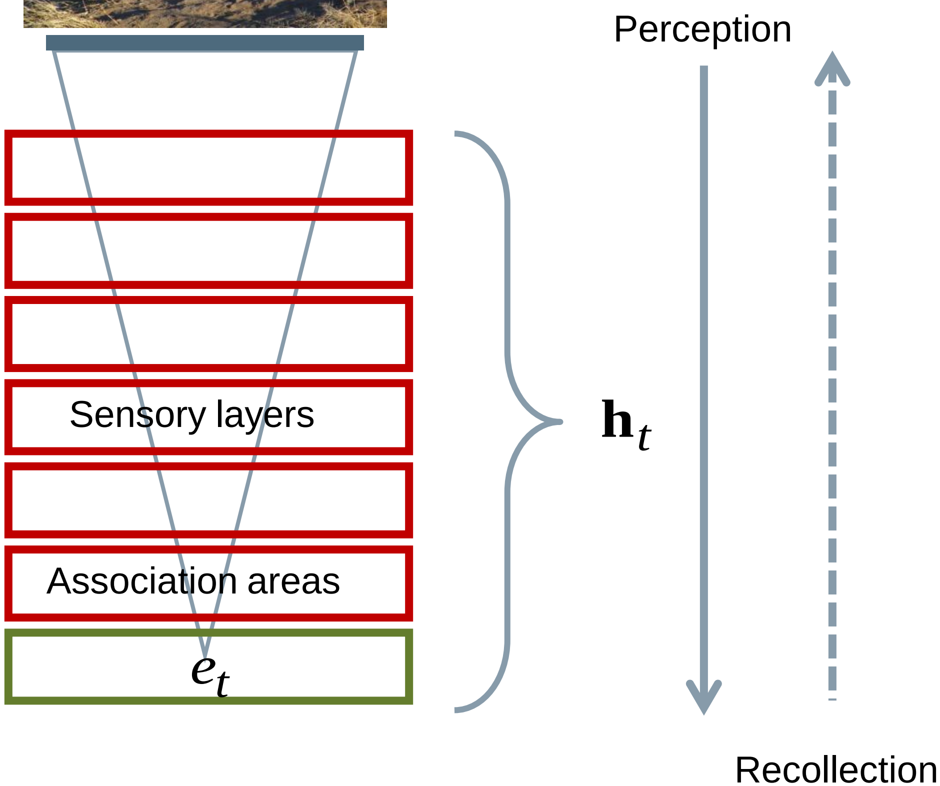


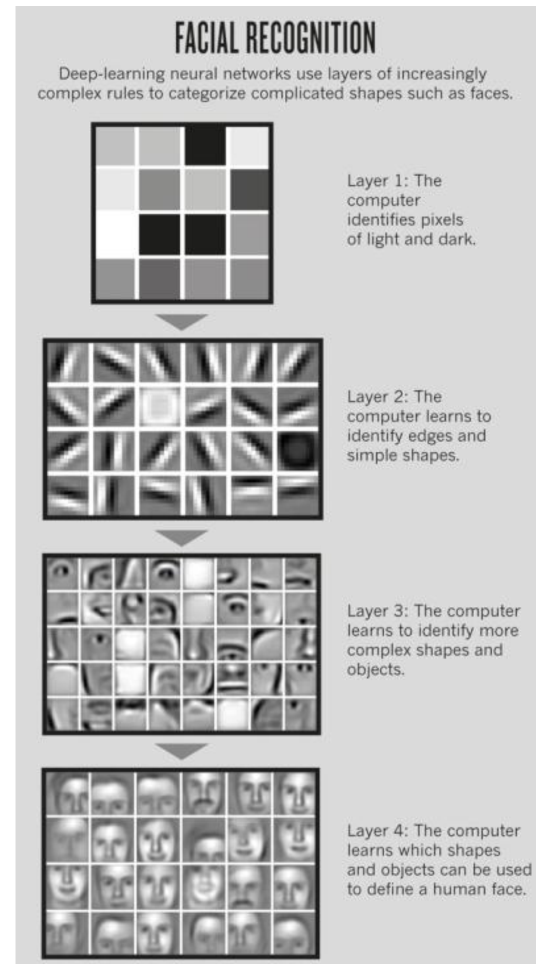
Sensory and other representations: looks, sounds, feels, …

Large overlap (general)

Little overlap (specific)

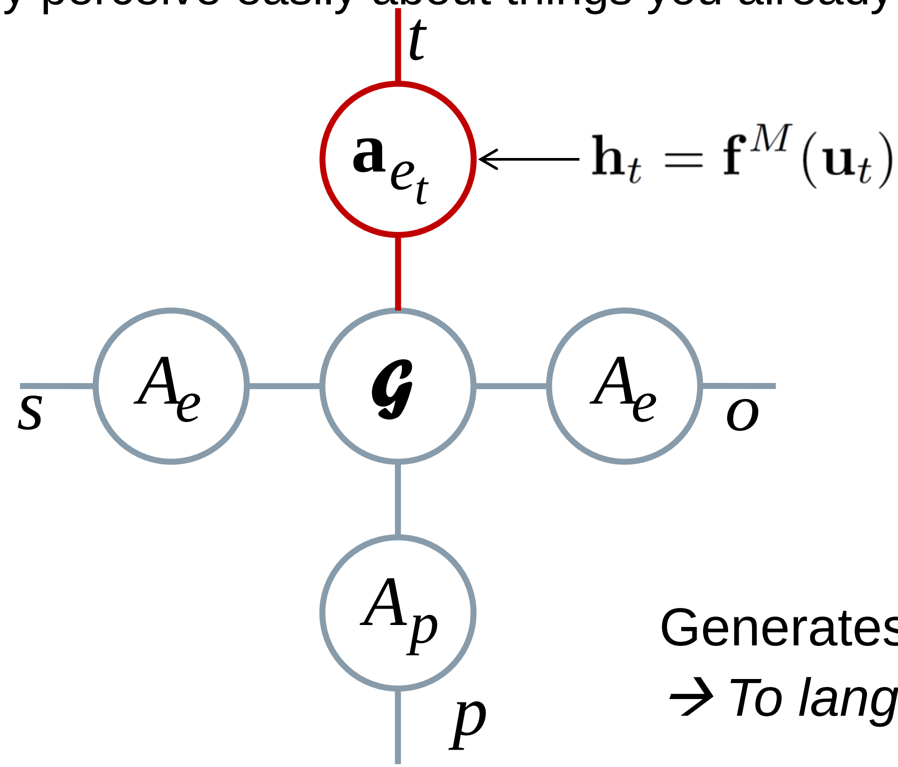
Index, new for each new memory





Declarative Decoding of Episodic Memory

- Recollection (external stage, declarative, symbolic): cortical processes operate on the output of the internal stage to reinstate the conscious experience of the episode (autonoetic consciousness)
- Episodic memory enters by the senses! (Can language communication directly generate episodic and semantic memories?)
- One can only perceive easily about things you already have traces



•From sensory layers

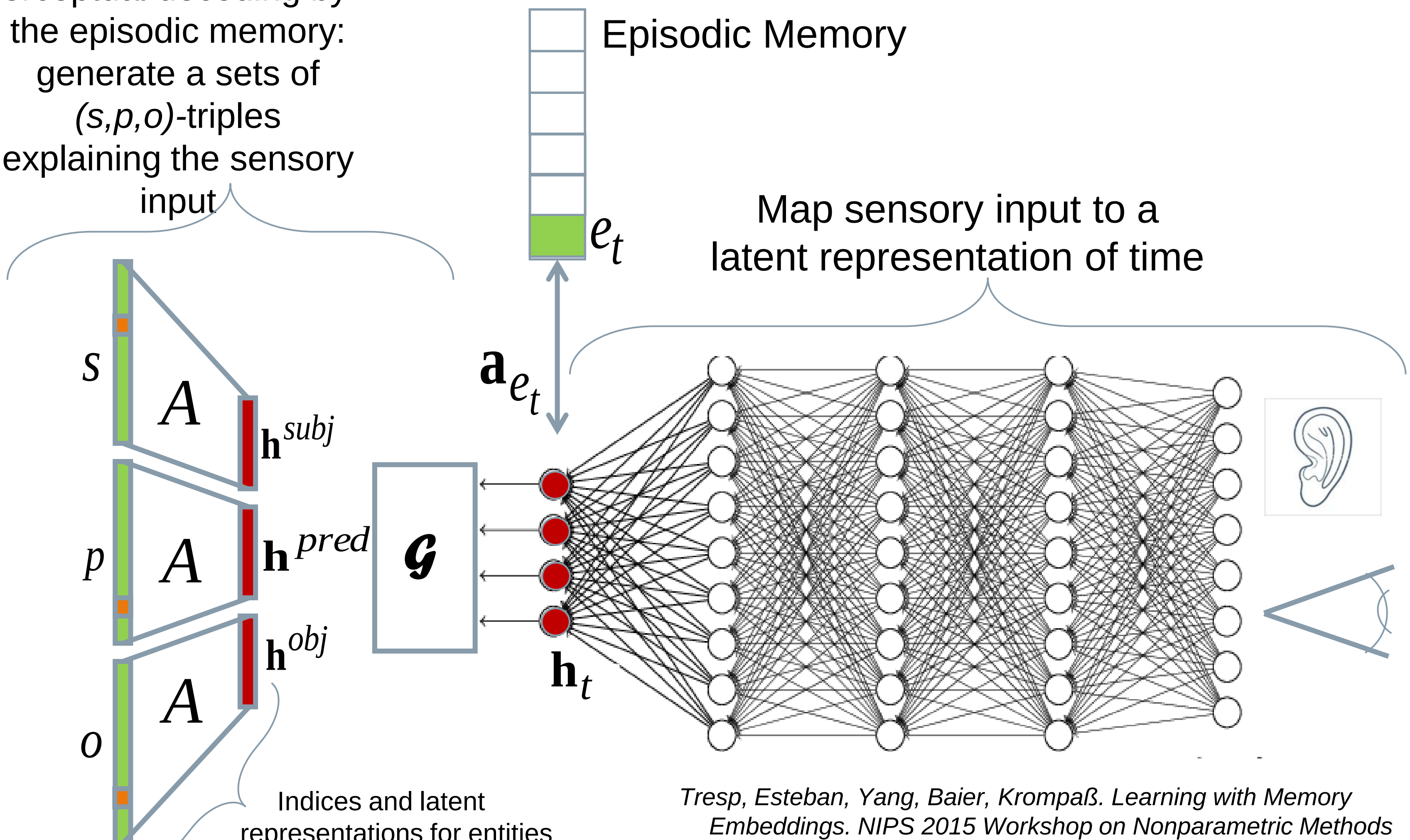
•From episodic memory

Generates: *(s,p,o), (s,p,o), ….*

→ *To language (Broca's area)???*

All-In-One Hypothesis: Perception, Episodic Memory and Semantic Memory all Use the Same Functional Brain Modules

Perceptual decoding by the episodic memory: generate a sets of *(s,p,o)*-triples explaining the sensory input

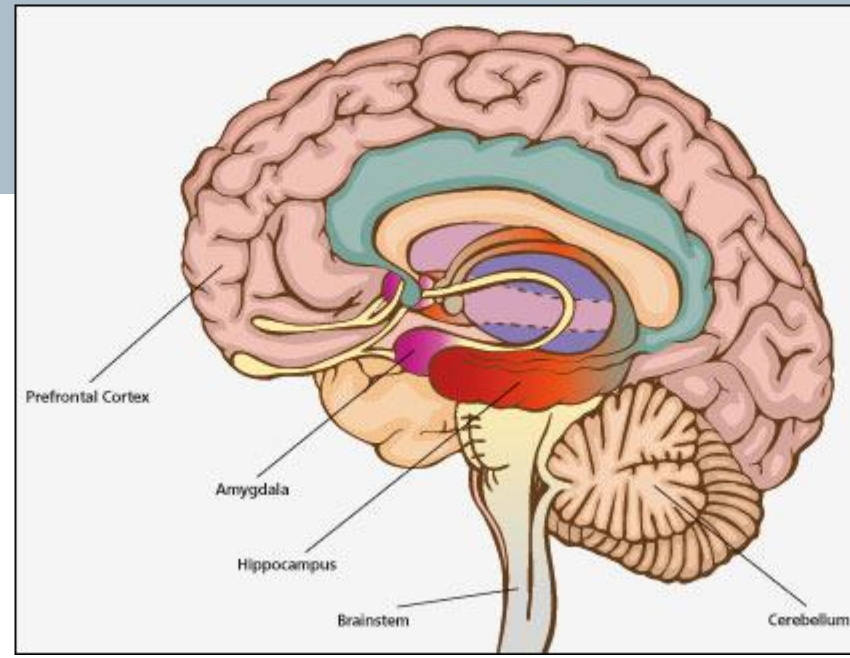


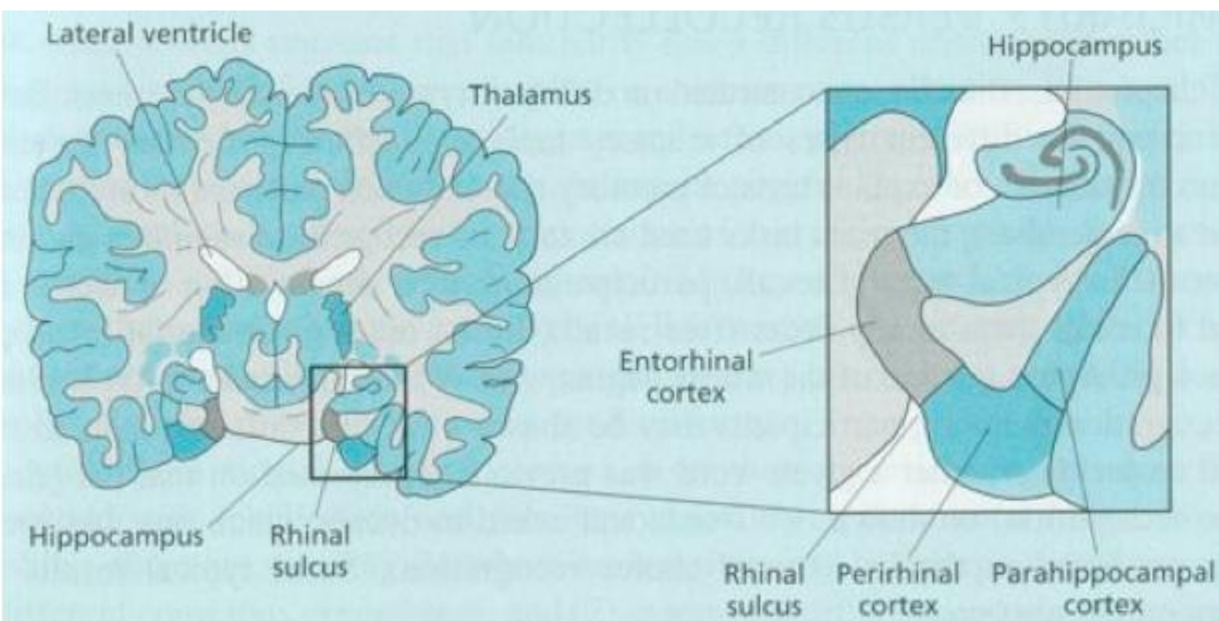
Indices and latent representations for entities and predicates

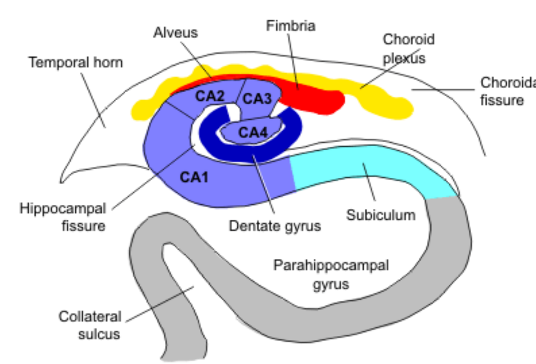
Tresp, Esteban, Yang, Baier, Krompaß. Learning with Memory Embeddings. NIPS 2015 Workshop on Nonparametric Methods
http://www.dbs.ifi.lmu.de/~tresp/papers/LearnMemEmb.pdf

Hippocampus and MTL play Significant Roles in the Generation of New Declarative Memories

- *Hippocampus-dependent declarative memory*
- New memories are formed in the hippocampus/MTL
  - Neurogenesis has been established in the **dentate gyrus** (part of the hippocampal formation) which is thought to contribute to the formation of **new episodic memories**
- Forming representations for new (significant)
  - **Episodes (time cells) (often)**
  - **Places (place cells) (often)**
  - **Entities (rare)**

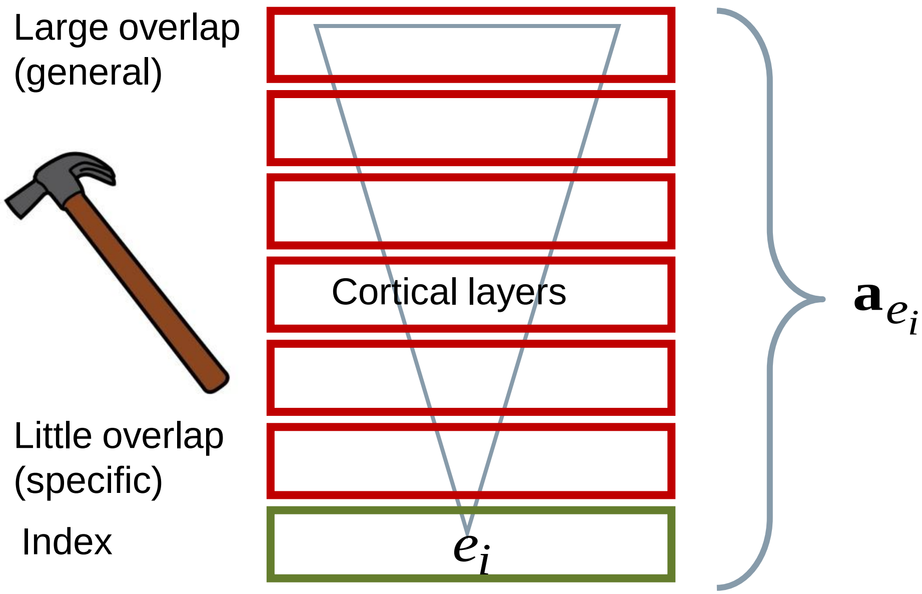




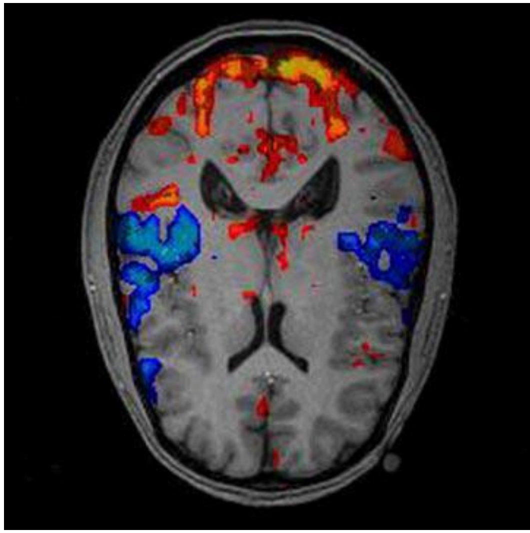


Representation of an Entity, a Concept, a Predicate

Hypothesis: in the same way that $\mathbf{h}_t$ / $\mathbf{a}_t$ represents the perception at time *t*, an entity e_i has a latent representation $\mathbf{a}_i$



Max, hammer, Munich, … …

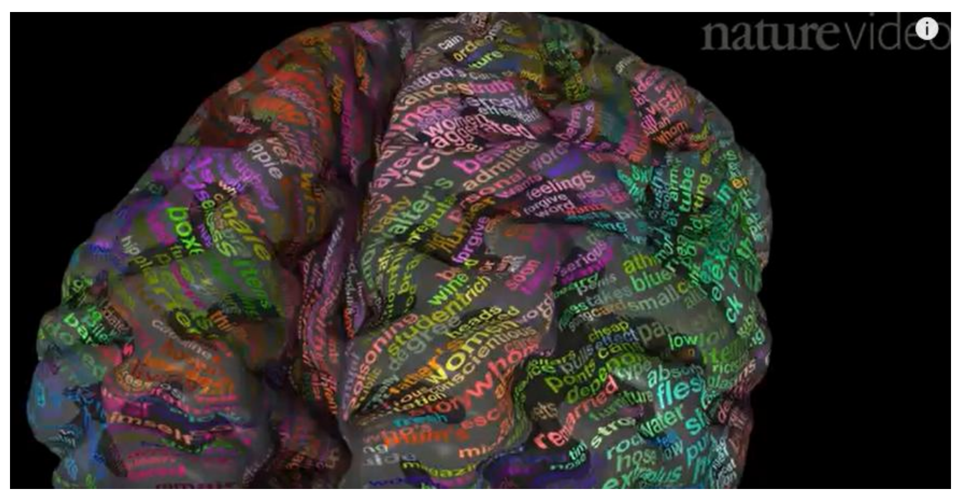


Locality of Representations for Concepts

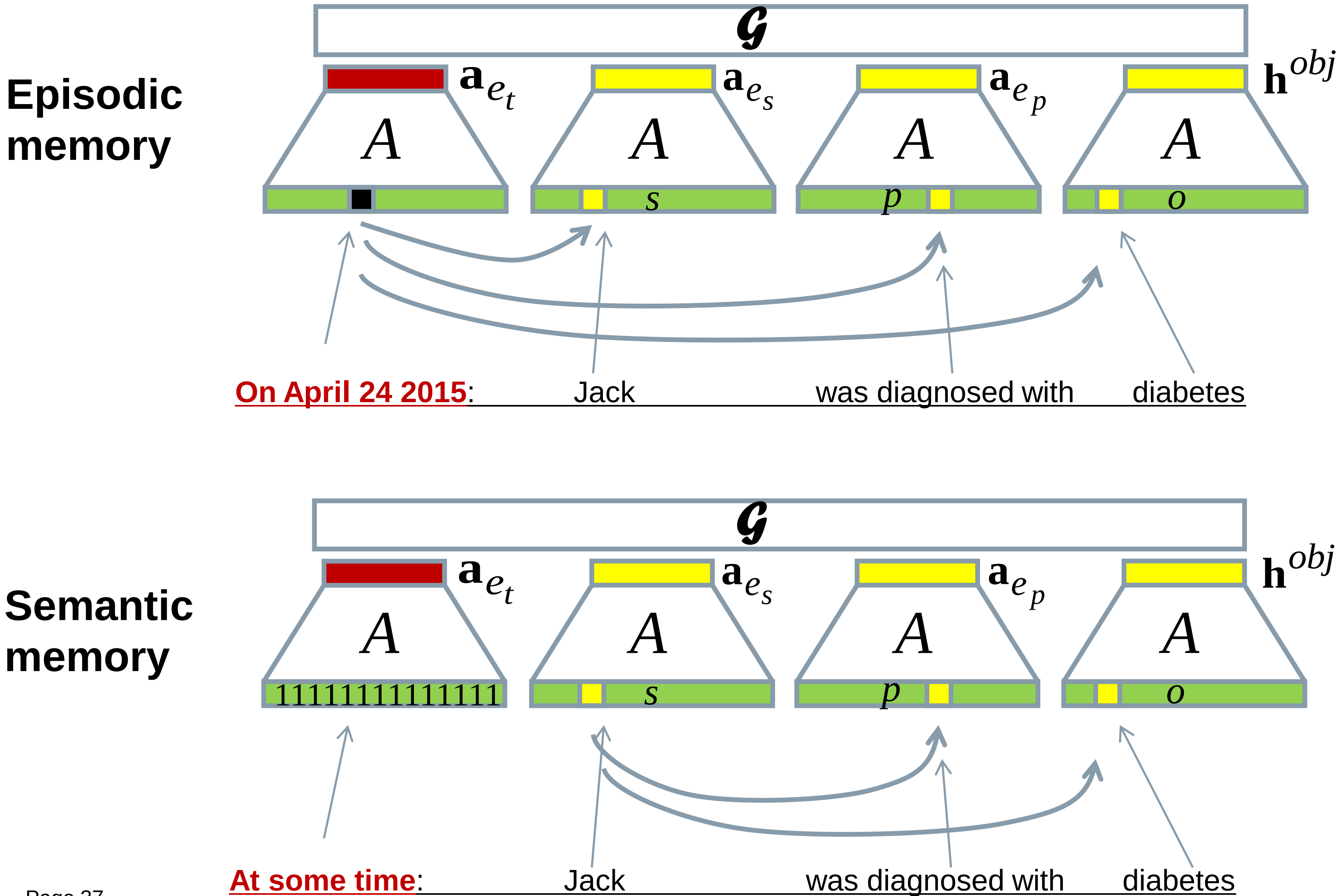
- Medial temporal lobe (MTL) neurons that are selectively activated by strikingly different **pictures of given individuals, landmarks or objects and in some cases even by letter strings with their names**
- “**Jennifer Aniston**“, „Halle Berry“ … concept cells

Quiroga, Reddy, Kreiman, Koch, Fried. Invariant visual representation by single neurons in the human brain. Nature, 2005

Huth, de Heer, Griffiths, Theunissen, Gallant. Natural speech reveals the semantic maps that tile human cerebral cortex. Nature, 2016.



Semantic Memory is Marginalized Episodic Memory



Semantic Memory from Episodic Memory?

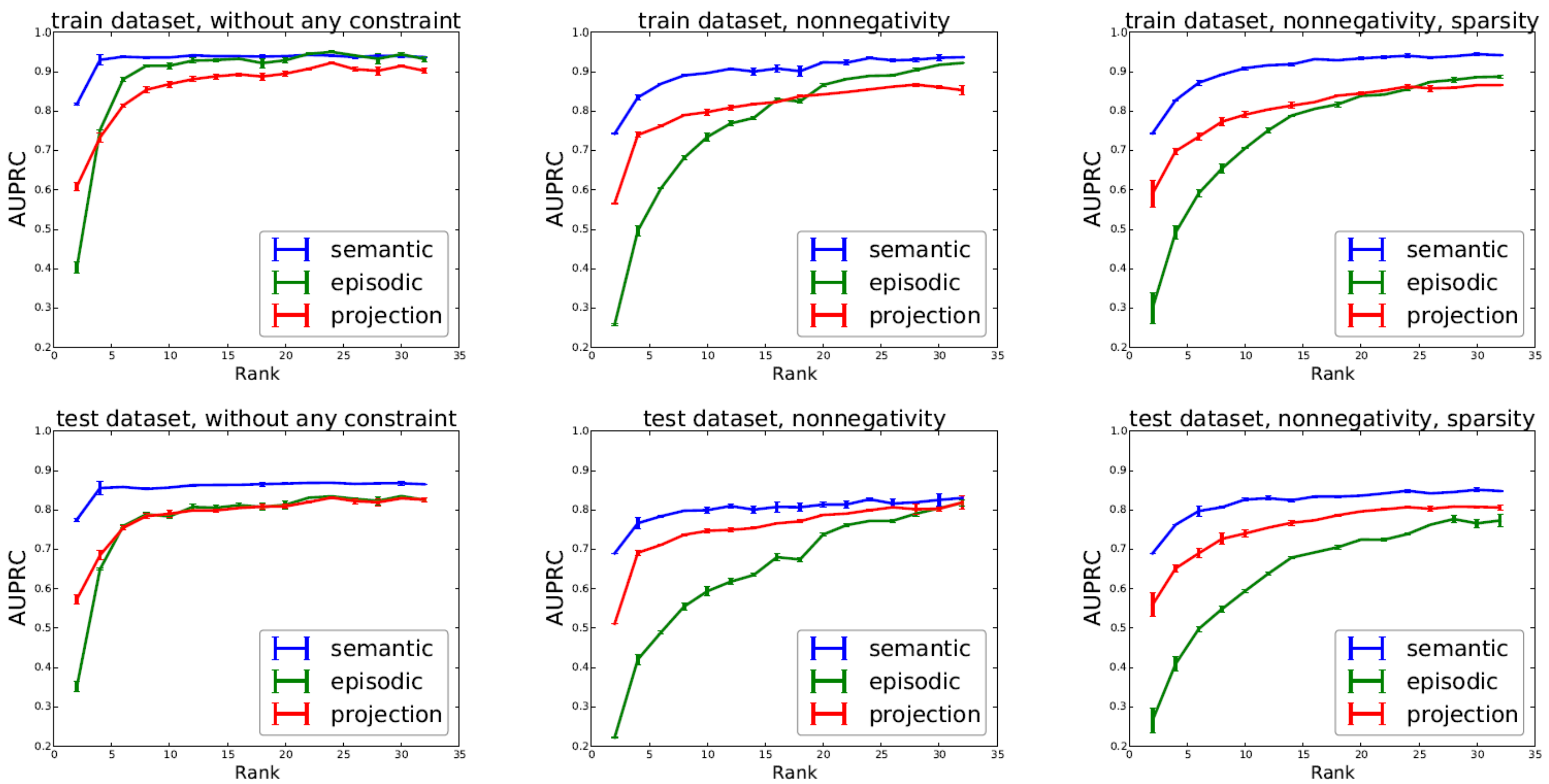


Fig. 3. AUPRC scores of the training and testing data sets for different model settings as a function of the rank.

Tresp, Ma, Baier, Yang. Embedding Learning for Declarative Memories. ESWC 2017

Integrated Conflict Early Warning System (ICEWS)

- The Integrated Conflict Early Warning System (ICEWS) data set that describes interactions between nations over several years
- An example entry could be (Turkey, Syria, Fight, 12/25/2014)
- A number of events have starting and end date

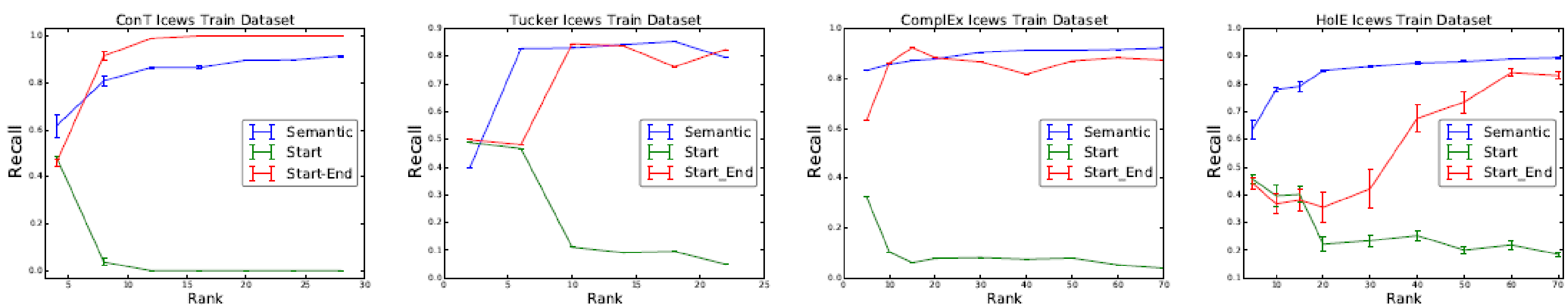


Figure 1: Recall scores vs. rank for the episodic-to-semantic projection on the ICEWS dataset with two different projection methods.

Memory Consolidation in the Real World

- No rapid eye movement!

Integrated Intelligence

Integrated Intelligence

Perception

(understand)

Integrated Intelligence

Episodic Memory

(remember)

Perception

(understand)

Integrated Intelligence

Semantic Memory

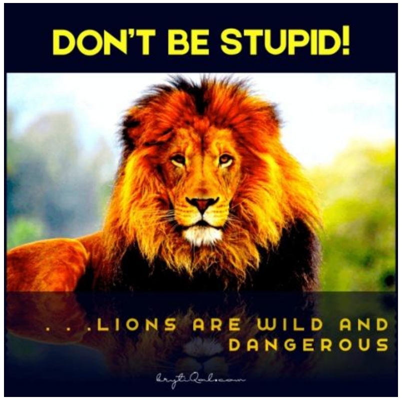
(know)

Episodic Memory

(remember)

Perception

(understand)



Integrated Intelligence

Semantic Memory

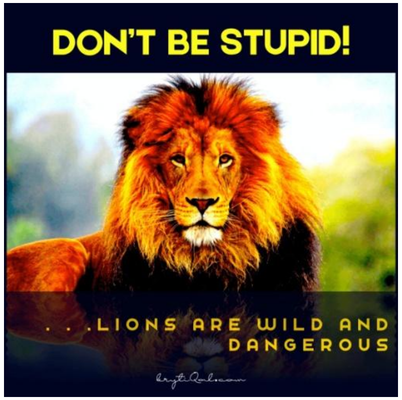
(know)

Episodic Memory

(remember)

Perception

(understand)



Integrated Intelligence

Declarative

Working Memory

- **Prediction**
- **Decision Making, Reinforcement Learning**
- **Planning**

Semantic Memory

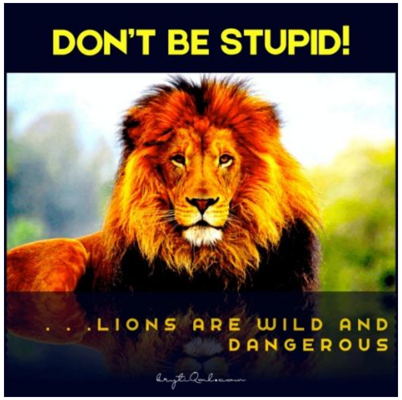
(know)

Episodic Memory

(remember)

Perception

(understand)



Integrated Intelligence

Working Memory
- **Prediction**
- **Decision Making, Reinforcement Learning**
- **Planning**

Semantic Memory

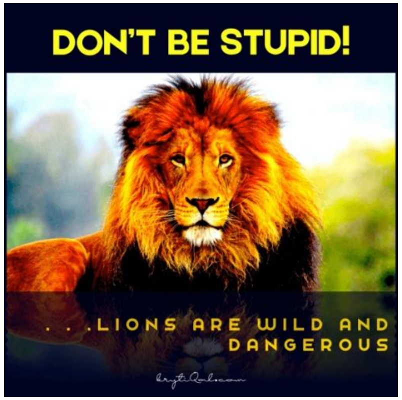
(know)

Episodic Memory

(remember)

Perception

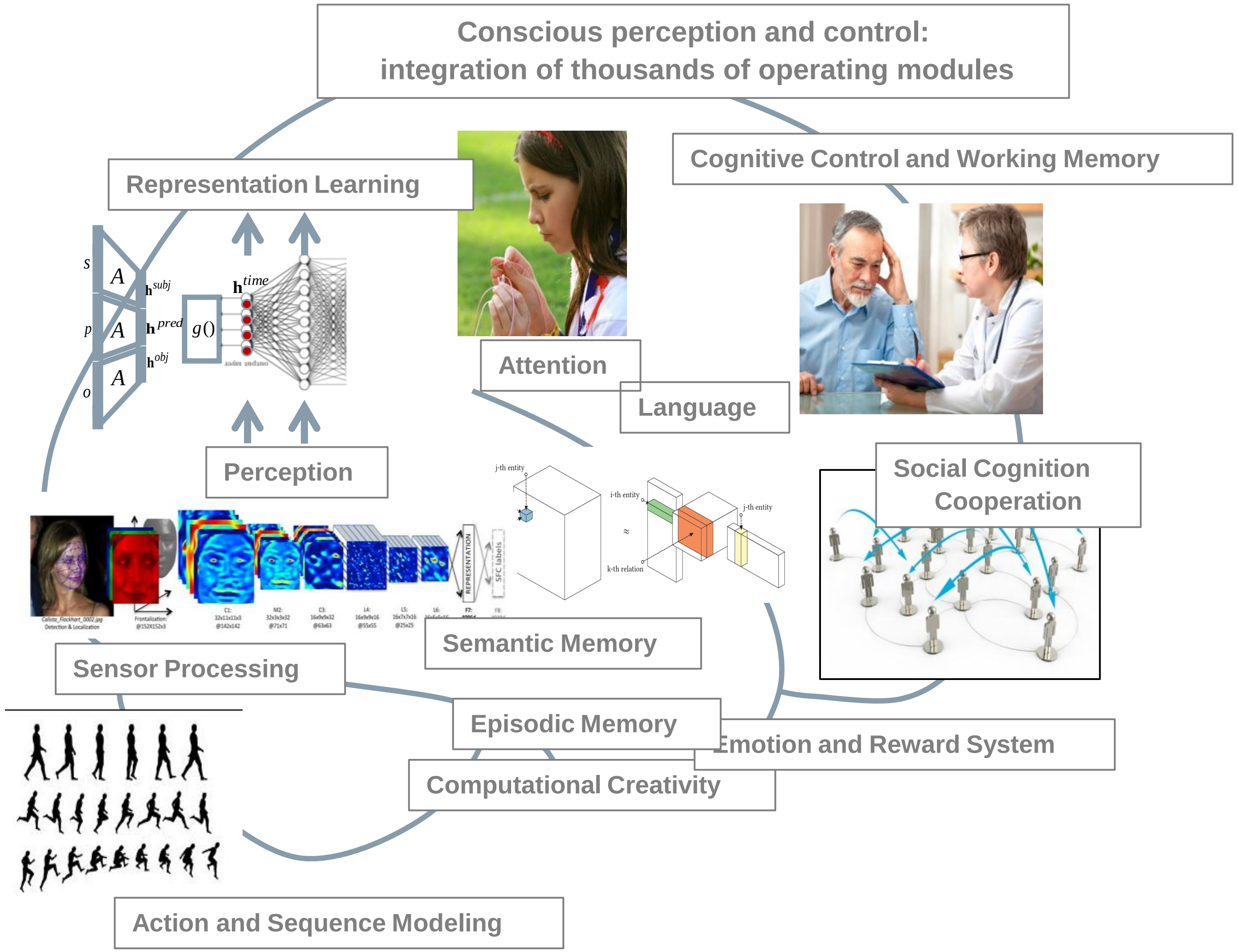
(understand)



AI is Many Things

Deep Learning with a Cognitive Perspective

Coordination of thousands of modules



Conclusions

- We see a lot of potential in the combination of knowledge graphs with machine learning
- A knowledge graph prior can be the basis for understanding sensory inputs, processed with deep learning
- Semantic and episodic knowledge graphs can be combined with deep learning in medical decision support
- The combination of knowledge graphs with machine learning might even be of interest to cognition:
  - We described a detailed mathematical models for the complete processing pipeline from sensory input and its semantic decoding, i.e., perception, to the formation of episodic and semantic memories and their **declarative** semantic decoding.

Sensory Memory	Sensor Data Processing (Spark Mini Batch)
Episodic Memory	Event Processing
Semantic Memory	Context Memory
Working Memory	Prediction, planning, decision making, reasoning

- *Tresp et al.. Learning with Memory Embeddings. NIPS 2015 Workshop on Nonparametric Methods for Large Scale Representation Learning*
- *Tresp et al.. The Tensor Memory Hypothesis. NIPS 2016 Workshop on Representation Learning in Artificial and Biological Neural Networks (MLINI 2016)*