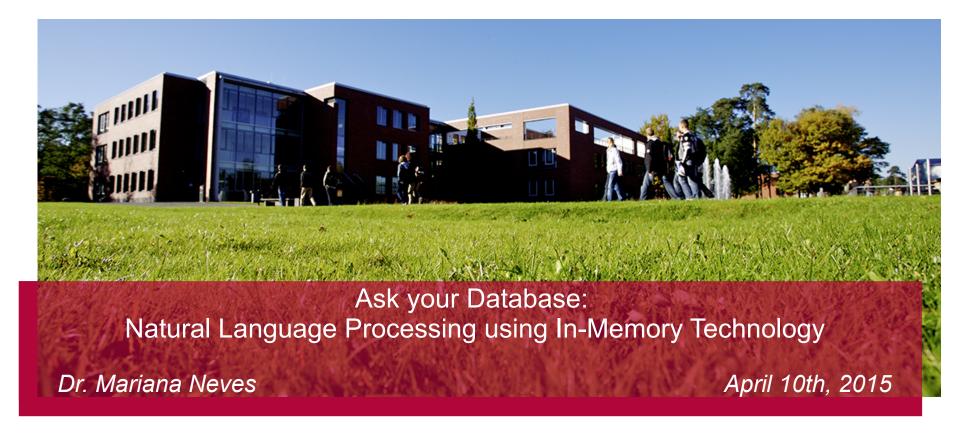
Enterprise Platform and Integration Concepts

Master Project Summer Term 2015



IT Systems Engineering | Universität Potsdam





Question Answering in Biomedicine

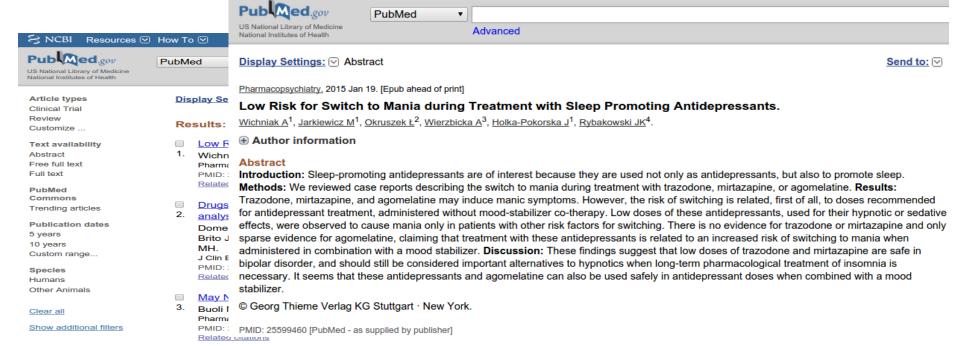


What disease is mirtazapine predominantly used for?











Question Answering in Biomedicine

What disease is mirtazapine predominantly used for?



major depression

"The 10 most commonly prescribed antidepressant drugs (citalopram hydrobromide (selective serotonin reuptake inhibitor), fluoxetine hydrochloride (selective serotonin reuptake inhibitor), amitriptyline hydrochloride (tricyclic antidepressant), paroxetine hydrochloride (selective serotonin reuptake inhibitor), venlafaxine hydrochloride (other), sertraline hydrochloride (selective serotonin reuptake inhibitor), mirtazapine (other), lofepramine (tricyclic antidepressant), and escitalopram (selective serotonin reuptake inhibitor)) comprised 93.6% (n=1) of all antidepressant prescriptions." (PMID 21810886)

"mirtazapine will make it the first-choice drug in depressive patients with gastric ulcers." (PMID 19034656)

"If antidepressants are used to treat insomnia, sedating ones should be preferred over activating agents such as serotonin reuptake inhibitors. In general, drugs lacking strong cholinergic activity should be preferred. Drugs blocking serotonin 5-HT2A or 5-HT2C receptors should be preferred over those whose sedative property is caused by histamine receptor blockade only. The dose should be as low as possible (e.g. as an initial dose: doxepin 25 mg, mirtazapine 15 mg, trazodone 50 mg, trimipramine 25 mg). Regarding the lack of substantial data allowing for evidence-based recommendations, we are facing a clear need for well designed, long-term, comparative studies to further define the role of antidepressants versus other agents in the management of insomnia." (PMID 19016570)





Time taken:2990ms What disease is mirtazapine predominantly used for Question: * ask Related Questions No related question! You asked: What disease is mirtazapine predominantly used for Clustered Answers Classified Answers Ranked Answers [mirtazapine] Serotonin-potentiating non-SSRIs, such as venlafaxine, nefazodone, trazodone and mirtazapine, have been evaluated in PTSD only in open-label and case studies..[Gregory:2004]{Human,Guideline} [predominantly] . Bupropion (amfebutamone), a predominantly noradrenergic reuptake inhibitor, was ineffective in PTSD in an openlabel study..[Gregory:2004]{Human,Guideline} [disorder] Post-traumatic stress disorder (PTSD) is a highly prevalent (7.8% lifetime rate) anxiety disorder with impairment in daily functioning, frequent suicidal behaviour and high rates of co-morbidity..[Gregory:2004]{Human,Guideline} [used] They should be avoided or used only short term because of potential depressogenic effects, and the possibility that they may promote or worsen PTSD..[Gregory:2004](Human, Guideline)



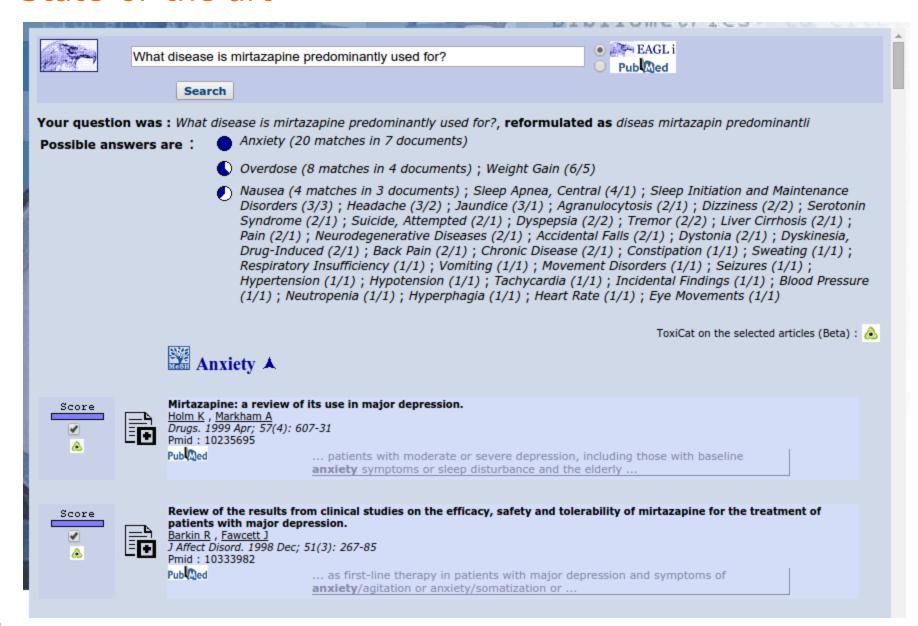




Time taken:2990ms						
Question:*	What disease is	mirtazapine predom	inantly used for		ask	
Related Questions						
No related question!						
You asked:What disease is mirtazapine predominantly used for						
Clustered Answers	Ranked Answers	Classified Answers				
 Post-traumatic stress disorder (PTSD) is a highly prevalent (7.8% lifetime rate) anxiety disorder with impairment in daily functioning, frequent suicidal behaviour and high rates of co-morbidity[Gregory:2004]{Human,Guideline} Serotonin-potentiating non-SSRIs, such as venlafaxine, nefazodone, trazodone and mirtazapine, have been evaluated in PTSD only in open-label and case studies[Gregory:2004]{Human,Guideline} Bupropion (amfebutamone), a predominantly noradrenergic reuptake inhibitor, was ineffective in PTSD in an open-label study[Gregory:2004] {Human,Guideline} They should be avoided or used only short term because of potential depressogenic effects, and the possibility that they may promote or worsen PTSD[Gregory:2004]{Human,Guideline} 						













What disease is mirtazapine predominantly used for?

Ask

Advanced Search

100 %

Answers: 43	Language: English	Expected question type: factoid	Expected medical type: treatment	Rate answers: Is it appropriate or not?	
1: <u>Answer</u> fro mirtazapine	m dailymed.nlm.nih.g tablet	Jov		4	×
	m www.drugs.com ng mirtazapine .			4	×
3: Answer from www.drugs.com doctor may prescribe Mirtazapine for				4	×
	m dailymed.nlm.nih.g to mirtazapine or an	jov y of the ingredients in mirtazapine ta	iblets.	4	×
	m www.drugs.com escribed Mirtazapine	e for a		4	×
	m www.ncbi.nlm.nih.q for treatment of depre	gov ession and comorbidities in Alzheime	r disease.	4	×
	m www.drugs.com in your blood.			4	×
	m www.drugs.com tablet, film coated				×

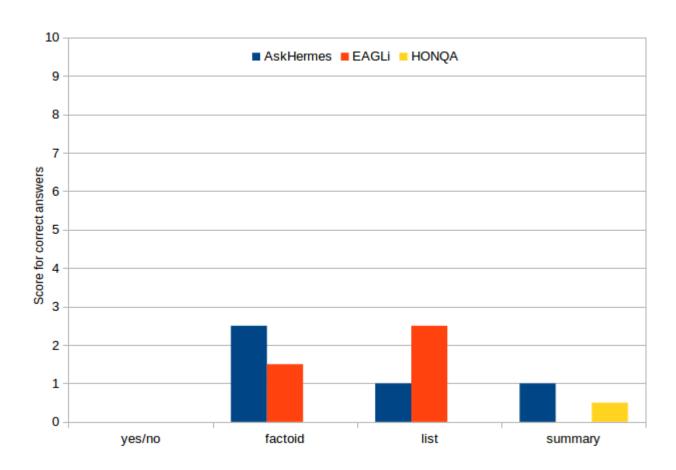










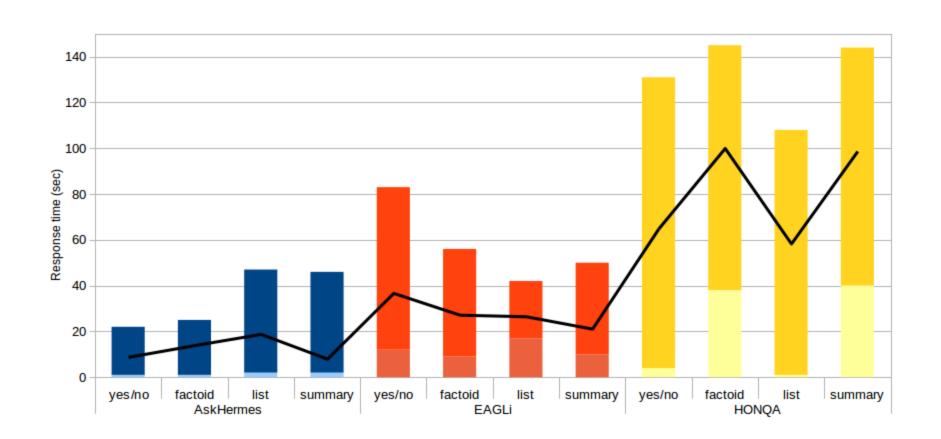








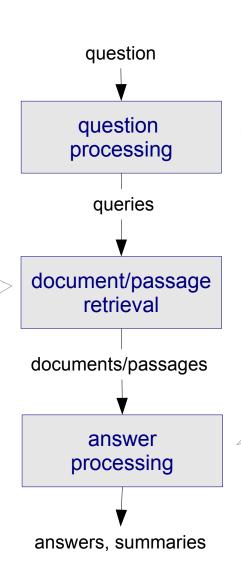






Architecture of QA systems

tokenization,
part-of-speech tagging,
chunking
syntatic parsing,
named-entity recognition,
document retrieval,
passage retrieval,
passages ranking,



tokenization,
part-of-speech tagging,
chunking
syntatic parsing,
term expansion,
named-entity recognition,

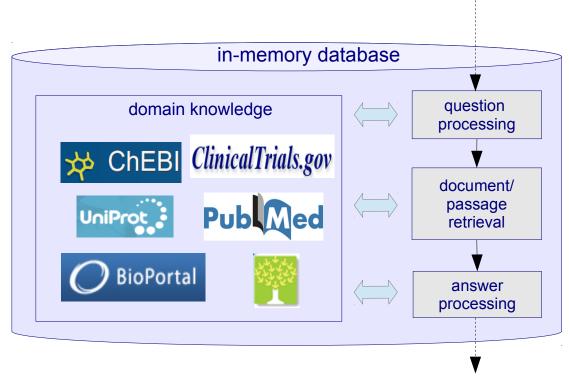
answer extraction, summarization, answer validation

. . .



QA architecture in IMDB





Answer: "major depression"

Summary: "Mirtazapine is predominantly used in the treatment of major depression." **Passages:** "second-generation antidepressants (selective serotonin reuptake inhibitors, nefazodone, venlafaxine, and <u>mirtazapine</u>) in participants younger than 19 years with MDD, OCD, or non-OCD anxiety disorders." (PMID 17440145)



NLP in SAP HANA

- Built-in features
 - Support to many languages
 - Document indexing
 - Fuzzy search
 - Sentence splitting, tokenization, part-of-speech tagging
 - Named entity recognition (dictionaries and/or rules)
 - Sentiments, negations(?), speculations(?)



Our QA prototype



A challenge on large-scale biomedical semantic indexing and question answering

System Name	→ Mean precision →	Recall →	F-Measur	e - MAP	• GMAP ▼	<u>.</u>	
Current Submission	0.0195	0.1493	0.0323	0.0497	7 0.0010		
HPI-S1	0.0097	0.0746	0.0163	0.0482	2 0.0004	_	
Wishart-S2	System Name	→ Mean p	recision -	Recall +	F-Measure -	MAP +	GMAP→
Wishart-S3	Current Submission	0.02	219	0.1568	0.0356	0.0714	0.0009
Wishart-S1	Wishart-S3	0.04	117	0.1268	0.0522	0.0574	0.0002
Top 100 Base	Wishart-S4	0.04	117	0.1268	0.0522	0.0574	0.0002
Top 50 Basel	Wishart-S1	0.05	501	0.1264	0.0516	0.0522	0.0002
_	Wishart-S2	0.05	501	0.1264	0.0516	0.0522	0.0002
_	HPI-S2	0.01	07	0.0857	0.0172	0.0517	0.0002
	Top 100 Baseline	0.06	668	0.0789	0.0577	0.0486	0.0002
_	Top 50 Baseline	0.06	586	0.0777	0.0596	0.0480	0.0002



Our QA prototype



A challenge on large-scale biomedical semantic indexing and question answering

```
"body": "Which extra thyroid tissues have thyrotropin (TSH) receptors?",
        "concepts": [
            "http://www.nlm.nih.gov/cgi/mesh/2012/MB cgi?field=uid&exact=Find+Exact+Term&term=D011989".
            "http://amigo.geneontology.org/cgi-bin/amigo/term_details?term=0004996",
            "http://www.biosemantics.org/jochem#4250044"
        ],
        "documents": [
            "http://www.ncbi.nlm.nih.gov/pubmed/22517745",
            "http://www.ncbi.nlm.nih.gov/pubmed/22496347",
            "http://www.ncbi.nlm.nih.gov/pubmed/22399514",
            "http://www.ncbi.nlm.nih.gov/pubmed/22289392",
            "http://www.ncbi.nlm.nih.gov/pubmed/21956421"
        ],
        "exact answer": [
            "adipose tissue", "fibrotic tissue"
        "id": "513f45abbee46bd34c000013",
        "ideal_answer": ["TSH receptors are expressed also in extrathyroid tissues. TSH receptors seem
to be functional. Extrathyroid tissues include fibrobasts of the orbit and adipose tissue\nThe
principal tissues with TSH receptors are:\nadippose tissue\n orbital fibrotic tissue"l.
        "snippets": [
                "beginSection": "sections.0",
                "document": "http://www.ncbi.nlm.nih.gov/pubmed/22517745",
                "endSection": "sections.0",
                "offsetInBeginSection": 1217,
                "offsetInEndSection": 1336,
                "text": "GD orbital fibroblasts, which comprise a mixture of CD34(+) and CD34(-) cells,
express much lower levels of Tg and TSHR"
                "beginSection": "sections.0",
                "document": "http://www.ncbi.nlm.nih.gov/pubmed/22517745".
                "endSection": "sections.0",
                "offsetInBeginSection": 552,
                "offsetInEndSection": 731.
                "text": "Previously, we found that CD34(+) progenitor cells, known as fibrocytes,
express functional TSHR, infiltrate the orbit, and comprise a large subset of orbital fibroblasts in
TAO. "
```

question processing



Sentence	split	ttina
0011100	٠٠٠٠	

What disease is mirtazapine predominantly used for?

Tokenization

What disease is mirtazapine predominantly used for

Part-of-speech tagging

What/WP disease/NN is/VBZ mirtazapine/NN predominantly/RB ..

Chunking

 $\text{[What disease]}_{\text{\tiny NP}} \text{ [is]}_{\text{\tiny VP}} \text{ [mirtazapine]}_{\text{\tiny NP}} \text{ [predominantly used]}_{\text{\tiny VP}}$

Entity recognition

mirtazapine/CHEMICAL_CHEBI:6950

Term expansion

disease [disorder; syndrome] mirtazapine [remeron] ...

Semantic role labeling

treat(agent:mirtazapine; theme:disease)

Expected answer

[DISEASE]

Question type

factoid

Query construction

disease [disorder; syndrome] AND mirtazapine [remeron] ...

(Analysis powered by Stanford parser, Genia tagger, becas, BioPortal)

document/passage retrieval



Sentence splitting

mirtazapine will make it the first-choice drug in depressive patients with gastric ulcers.

METHODS: Antidepressant therapy with 15 to 45 mg/d of mirtazapine (n = 124) or 20 to 40 mg/d of paroxetine (n = 122)

Tokenization

METHODS: Antidepressant therapy with 15 to 45 mg / d of

Part-of-speech tagging

Patients/NNS 65/CD years/NNS or/CC older/JJR with/IN ...

Chunking

 $[patients 65 years]_{NP}$ or $[older]_{ADJP}$ $[with]_{PP}$ $[major depression]_{NP}$

Entity recognition

major depression/DISORDER_SNOMEDCT:370143000 mirtazapine/CHEMICAL_CHEBI:6950 gastric/FMA:Stomach

Semantic role labeling

treat (agent:mirtazapine; theme: depression)

Passage retrieval

METHODS: Antidepressant therapy with 15 to 45 mg/d of mirtazapine (n = 124) or 20 to 40 mg/d of paroxetine (n = 122)

Passages ranking

(0.8) METHODS: Antidepressant therapy with 15 to 45 mg/d of mirtazapine (n = 124) or 20 to 40 mg/d of paroxetine (n = 122)

(0.5) mirtazapine will make it the first-choice drug in depressive patients with gastric ulcers.

(Analysis powered by Stanford parser, Genia tagger and becas)

answer processing



Answer extraction

SRL-Question: treat(agent:mirtazapine; theme:disease)

SRL-Passage: treat (agent:mirtazapine; theme: depression)

Expected answer: [disease]

NER: major depression/DISORDER_SNOMEDCT:370143000

 $\rightarrow \rightarrow \rightarrow$ major depression

Summarization

- Mirtazapine will make it the first-choice drug in depressive patients with gastric ulcers.

- METHODS: Antidepressant therapy with 15 to 45 mg/d of mirtazapine (n = 124) or 20 to 40 mg/d of paroxetine (n = 122)

 $\rightarrow \rightarrow \rightarrow$ "Antidepressant therapy with mirtazapine will make it the first-choice drug in depressive patients"

Answer validation

- major depression
- "Antidepressant therapy with mirtazapine will make it the first-choice drug in depressive patients"

Mirtazapine: a review of its use in major depression. www.ncbi.nlm.nih.gov/pubmed/10235695 ▼

by KJ Holm - 1999 - Cited by 150 - Related articles

Mirtazapine: a review of its use in **major depression**. ... has **predominantly** been evaluated in the treatment of **major depression**. ... or **severe depression**, including those with baseline anxiety **symptoms** or sleep disturbance and the elderly.

Mirtazapine Uses, Dosage & Side Effects - Drugs.com www.drugs.com → Drugs A to Z ▼

Mirtazapine (Remeron) is an antidepressant used to treat major depressive disorder. Includes mirtazapine side effects, interactions and indications.



Goals

- Implement new NLP functionalities in SAP HANA database
- Adapt current NLP features in SAP HANA to the biomedical domain
- Participate in the development of a QA system for the biomedical domain
- Evaluate the application on the BioASQ dataset



- Groups
 - Natural Language Processing in IMDB (3 students)
 - Question Answering for Biomedicine (6 students)



Group 1-NLP:

Natural Language Processing in IMDB

What disease is mirtazapine predominantly used for?

Tokenization

What disease is mirtazapine predominantly used for

Part-of-speech tagging

What/WP disease/NN is/VBZ mirtazapine/NN predominantly/RB ...

Chunking [What disease]_{NP} [is]_{VP} [mirtazapine]_{NP} [predominantly used]_{VP}

Semantic role labeling treat(agent:mirtazapine; theme:disease)



Group 1-NLP:

Natural Language Processing in IMDB

- Tokenization, Part-of-speech tagging, Chunking, Semantic role labeling
- Supervised learning (Support Vector Machines): Predictive Analysis Library (R, C++)
- Corpora: GENIA, TweetNLP, CoNLL-2000, CoNLL-2005, BioProp
- Integration into HANA full text indexing



Group 2 - QA: Question Answering for Biomedicine

QA-1: Question processing

What disease is mirtazapine predominantly used for?

Expected answer [DISEASE]

Question type factoid

Query construction

disease [disorder; syndrome] AND mirtazapine [remeron] ...



Question Answering for Biomedicine

- QA-1: Question processing
 - query construction, question type and target recognition
 - Output from the NLP group
 - SQL, stored procedures
 - Supervised learning (Support Vector Machines):
 Predictive Analysis Library (R, C++)
 - Corpus: BioASQ



Question Answering for Biomedicine

QA-2: Named-entity recognition

What disease is mirtazapine predominantly used for?

Entity recognition

mirtazapine/CHEMICAL CHEBI:6950

mirtazapine will make it the first-choice drug in depressive patients with gastric ulcers. METHODS: Antidepressant therapy with 15 to 45 mg/d of mirtazapine (n = 124) or 20 to 40 mg/d of paroxetine (n = 122)

Entity recognition

major depression/DISORDER_SNOMEDCT:370143000 mirtazapine/CHEMICAL_CHEBI:6950 gastric/FMA:Stomach



Group 2 - QA: Question Answering for Biomedicine

- QA-2: Named-entity recognition
 - integration of various resources
 - HANA's built-in NER features
 - Corpus/evaluation: various + BioASQ



Question Answering for Biomedicine

QA-3: Passage retrieval

Passage retrieval

METHODS: Antidepressant therapy with 15 to 45 mg/d of mirtazapine (n = 124) or 20 to 40 mg/d of paroxetine (n = 122)

Passages ranking

- (0.8) METHODS: Antidepressant therapy with 15 to 45 mg/d of mirtazapine (n = 124) or 20 to 40 mg/d of paroxetine (n = 122)
- (0.5) mirtazapine will make it the first-choice drug in depressive patients with gastric ulcers.



Question Answering for Biomedicine

- QA-3: Passage retrieval
 - Ranking passages/sentences
 - Calculate a score
 - PubMed indexed in Future SOC lab
 - Output from NLP and NER groups
 - SQL, stored procedures
 - Evaluation: BioASQ



Group 2 - QA: Question Answering for Biomedicine

QA-4: Answer processing

Answer extraction

SRL-Question: treat(agent:mirtazapine; theme:disease)
SRL-Passage: treat (agent:mirtazapine; theme: depression)
Expected answer: [disease]

NER: major depression/DISORDER_SNOMEDCT:370143000

 $\rightarrow \rightarrow \rightarrow$ major depression



Question Answering for Biomedicine

- QA-4: Answer processing
 - Answers for factoid/list/yes-no questions
 - Output from NER and NLP groups
 - SQL, stored procedures
 - Evaluation: BioASQ



Question Answering for Biomedicine

- QA-4: Answer processing
 - Answers for factoid/list/yes-no questions
 - Output from NER and NLP groups
 - SQL, stored procedures
 - Evaluation: BioASQ



Question Answering for Biomedicine

QA-5: Summarization

- Mirtazapine will make it the first-choice drug in depressive patients with gastric ulcers.

Summarization

- METHODS: Antidepressant therapy with 15 to 45 mg/d of mirtazapine (n = 124) or 20 to 40 mg/d of paroxetine (n = 122)

→→→ "Antidepressant therapy with mirtazapine will make it the first-choice drug in depressive patients"



Group 2 - QA: Question Answering for Biomedicine

- QA-5: Summarization
 - Answers for summary questions
 - SQL, stored procedures
 - Evaluation: BioASQ



Demo Web tool

- Input: question
- Output:
 - Answers (exact + summary)
 - Relevant passages (highlighted entities)
 - Feedback from users



Grading

Report/paper: 40%

- Commitment: 10%

- Implementation: 50%



- Web page: https://hpi.de//plattner/teaching/summerterm2015/masterproject2015.html
- Room: Campus D, E.0-1.2
- Mail with 3 top choices to "mariana.neves@hpi.de" until Mon, April 13th
- Meeting every 2 weeks (Mariana/Cindy)
- Mid-term presentation (June) and final presentation (July)



Contact: Villa, HPI Campus II, room V1.02



Dr. Mariana Neves mariana.neves@hpi.de



Cindy Fähnrich, Msc. cindy.faehnrich@hpi.de



Dr. Matthias Uflacker matthias.uflacker@hpi.de



- Mail with 3 top choices to "mariana.neves@hpi.de" until Mon, April 13th
 - NLP: Natural Language Processing in IMDB
 - QA-1: Question processing
 - QA-2: Named-entity recognition
 - QA-3: Passage retrieval
 - QA-4: Answer processing
 - QA-5: Summarization