# Problembeschreibungen

Die folgenden Probleme bestehen

1. Ein medizinischer Text muss vorverarbeitet werden
2. Medizinisches Fachvokabular muss in ein computerauswertbare Form überführt werden
3. Eine Wissensrepräsentation muss erstellt werden

# Hypothesen?

# Forschungsfragen

1. Wie kann ein medizinischer Text vorverarbeitet werden, so dass relevante Aussagen (Überschriften, Aufzählungen, Leerzeilen und Dialoge) automatisch erkannt werden?
2. Wie kann ein medizinisches Fachvokabular über depressive Erkrankungen automatisiert in eine maschinenlesbare Form überführt werden?
3. Wie kann eine Wissensrepräsentation über einen gegebenen Text maschinenlesbar erstellt werden?

# Forschungsziele

1. Wie kann ein medizinischer Text vorverarbeitet werden, so dass relevante Aussagen (Überschriften, Aufzählungen, Leerzeilen und Dialoge) automatisch erkannt werden?
   1. Aufbau und Struktur medizinischer Texte (Observation)
   2. Theoriebildung zur Vorverarbeitung medizinischer Texte (Theoriebildung)
   3. Implementierung zur Vorverarbeitung medizinischer Texte (Implementierung)
   4. Evaluation zur Vorverarbeitung medizinischer Texte (Evaluation)
2. Wie kann ein medizinisches Fachvokabular über depressive Erkrankungen automatisiert in eine maschinenlesbare Form überführt werden?
   1. Überblick über depressive Erkrankungen (Observation)
   2. Automatisierung durch NLP (Observation)
   3. Theoriebildung zur Überführung medizinischen Fachvokabulars in maschinenlesbare Form zur weiteren Verarbeitung durch NLP (Theoriebildung)
   4. Implementierung zur Überführung medizinischen Fachvokabulars in maschinenlesbare Form zur weiteren Verarbeitung durch NLP (Implementierung)
   5. Evaluation zur Überführung medizinischen Fachvokabulars in maschinenlesbare Form zur weiteren Verarbeitung durch NLP (Evaluation)
3. Wie kann eine Wissensrepräsentation über einen gegebenen Text maschinenlesbar erstellt werden?
   1. Wissensrepräsentation mittels RDF, RDFS (Observation)
   2. Theoriebildung zur Wissensrepräsentation (Theoriebildung)
   3. Implementierung zur Wissensrepräsentation (Implementierung)
   4. Evaluation zur Wissensrepräsentation (Evaluation)

## Beobachtung

State of the Art

## Theoriebildung

## Systementwicklung

## Experiment

* Manuelle Erstellung einer RDF/XML-Datei
* Vergleich von manuell und automatisch erstellter Datei und Auswertung

Lösungsansätze:

1. Import des MeSh-Vokabulars in spaCy
2. Entity-Linking
   1. Struktur und Format einer sinnvollen Wissensrepräsentation
   2. Nutzung der Grammatikanalyse von spaCy zur Erkennung von Zusammenhängen
   3. Export als RDF/XML
3. \begin{question}
4. Wie kann die MEsH Datenbank in spaCy importiert und von der NER-Komponente genutzt werden?
5. \end{question}
6. \begin{question}
7. Welche Entities werden mit der MEsH Datenbank gefunden? Wie unterstützen diese Entities den Aufbau der Wissensrepräsentation?
8. \end{question}
9. \begin{question}
10. Was muss in der Wissensdatenbank der Entity-Linker-Komponente von spaCy gespeichert werden?
11. \end{question}
12. \begin{question}
13. Wie kann der Text in einzelne Sätze gegliedert werden, so dass auch Überschriften, Leerzeilen und Aufzählungen berücksichtigt werden?
14. \end{question}
15. \begin{question}
16. Wie können Überschriften erkannt werden?
17. \end{question}
18. \begin{question}
19. Wie können Aufzählungen erkannt und einem Thema zugeordnet werden?
20. \end{question}
21. \begin{question}
22. Wie sieht eine sinnvolle Wissensrepräsentation aus?
23. \end{question}