# Aufgabe 1

## Analyse:

1. Das Dokument ist in utf-8 Kodierung, es enthält nicht-ASCII Symbole wie z.B. chinesische Hieroglyphen.
2. Außer Links enthält das Dokument diverse undefinierte Textausschnitte, welche vom Programm ignoriert werden sollen.
3. Unterschiedliche Linkformate:

* Allgemeine Links [[link]] und Piped links [[link|text]]. Bei allgemeinen Links ist Zielartikelname gleich dem gelinkten Text.
* „Internal links to an anchor“ ([[#See also]]), „internal links to a subpage“ ([[/example]]): Name der Seite kann nicht ermittelt werden, daher sollen diese Links ignoriert werden.
* Links mit „Pipe trick“ [[link|]] sollen als allgemeine Links behandelt werden

1. Geschachtelte Links: nur das interne Link soll verarbeitet werden (Verhalten der Wikipedia).
2. Links mit \n:

* Zielartikelnamen mit einem Zeilenumbruch sollen ignoriert werden.
* Zeilenumbruch im gelinkten Text mit Leerzeichen ersetzen (Verhalten der Wikipedia).

## Programmbeschreibung

Das Programm analysiert die Eingabedatei("data/wikilinks\_en.txt") symbolweise. Für das Finden von Anfang („[[„) und Ende („]]“) von Links wird ein vorheriges Symbol gespeichert. Wenn Linkanfang gefunden wurde, werden darauffolgende Symbole als Name der gelinkten Artikel interpretiert, bis ein Linkende bzw. Pipe („|“) gefunden wird. Falls ein Pipe gefunden wurde, werden darauffolgende Symbole als Linktext interpretiert, bis ein Linkende gefunden wird. Sonst wird wie in „[Analyse](#_Analyse:)“ beschrieben vorgegangen.

Die Ausgabe bestehend aus Paaren Zielartikelnamen und Linktext wird unter "data/Linktext.txt" gespeichert.

Für die Perfomanceanalyse wird Laufzeit des Programms angezeigt.