

Das ZIH (Zentrum für Informationsdienste und Hochleistungsrechnen) Service Desk erhält im Laufe der Jahre eine große Menge an E-Mails mit Fragen zu Themen wie z. B. Gruppenlaufwerken, E-Mail-Services oder auch Mailinglisten. Viele von diesen Mails beinhalten Fragen, welche sehr häufig gestellt werden, um über eine FAQ -Website vermieden werden könnten. Im Rahmen dieser Arbeit werden verschiedene Ansätze vorgestellt, häufig gestellte Fragen zu erkennen. Unter Verwendung des im letzten Jahr erschienenen BERT-Modells [5] wird hierbei die sogenannte Next Sentence Prediction verwendet, um Sätze mit thematischer bzw. semantischer Ähnlichkeit zu finden. Anhand dieser Sätze können Clustering-Verfahren angewendet werden, welche die Mails gruppieren. Die entstandenen Gruppen werden analysiert und bewertet, um FAQs aus ihnen zu formulieren.