

Generation of Meaningful SQL-Query Exercises Using Large Language Models and Knowledge Graphs

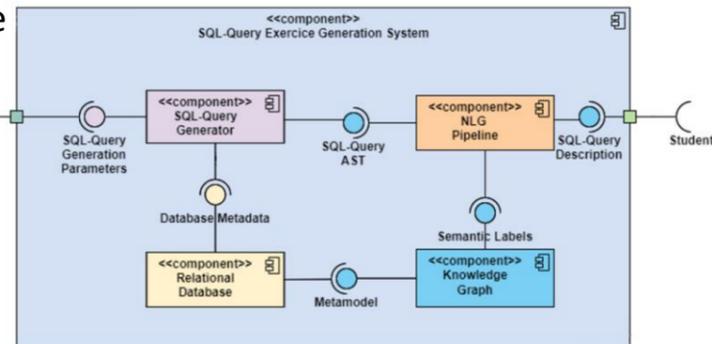
Paul Christ

Kurzzusammenfassung

Dieser Arbeit präsentiert ein System zur automatischen Generierung von SQL-Abfrage-Übungen, die einer Reihe von Eingabeparametern genügen und syntaktisch korrekte und semantisch sinnvolle SQL-Abfragen erzeugen. Eine SQL-Abfrage-Übung besteht aus einer SQL-Abfrage, die vor dem Studenten verborgen ist, und einer natürlichsprachlichen Beschreibung. Das Ziel der Übung ist es, eine äquivalente SQL-Abfrage auf der Grundlage der natürlichsprachlichen Beschreibung und des Schemas der zugrunde liegenden Datenbank zu reproduzieren.

System

Das vorgeschlagene System besteht aus vier zentralen Komponenten. Einem Wissensgraphen, einer

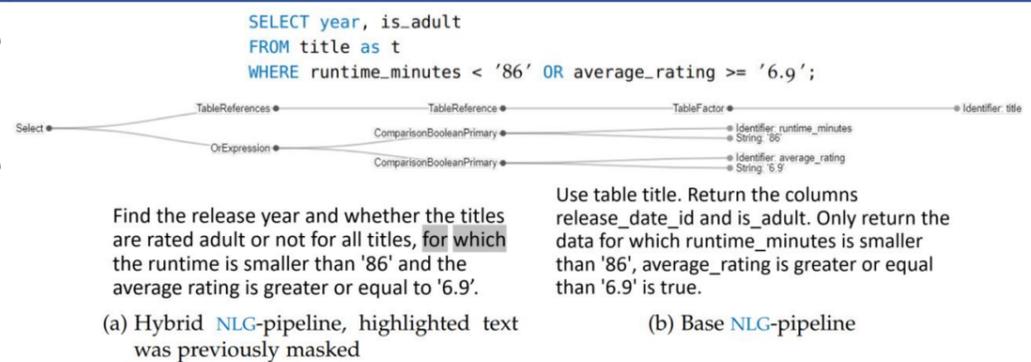


relationalen Datenbank, dem SQL-Abfragegenerator selbst und einer Pipeline zur Erzeugung natürlicher Sprache. Der Wissensgraph dient als Metamodell, um weitere Informationen für die NLG-Pipeline und die relationale Datenbank bereitzustellen.

Methodik

Aufgrund des breiten Spektrums relevanter Disziplinen, wie relationale Datenbanken, natürliche Sprachverarbeitung und Graphentheorie, wurden, falls möglich, vorhergehende Ergebnisse in das System integriert. Da zwischen den Teilergebnissen Abhängigkeiten bestehen, wird auf Alternativen zurückgegriffen, wenn eine Basiskomponente als unzureichend erachtet wird. Die Evaluierung des Systems wurde anhand von menschlichem Feedback durchgeführt. Die Datenerhebung wurde aus Kosten- und Zeitgründen zu einem einzigen Zeitpunkt durchgeführt und es erfolgte kein unmittelbarer Rückfluss von Informationen.

System-Output



Das System erzeugt SQL-Abfragen, zusammen mit einem abstrakten Syntaxbaum, der zur Erzeugung einer natürlichsprachlichen Beschreibung verwendet wird. Die NLG-Pipeline nutzt semantische Bezeichnungen, die durch den Wissensgraphen bereitgestellt werden und ein Sprachmodell um fehlende Textbausteine zu ergänzen.

Resultate

Ein Datensatz von 667 SQL-Abfragen, die vom System generiert wurden, wurde mit dem Crowdsourcing-Framework Amazon Mechanical Turk bewertet. Jede Aufgabe wurde von drei Crowdworkern bewertet. Alle teilnehmenden Crowdworker gaben an, grundlegende SQL-Kenntnisse zu besitzen. Die Umfrage besteht aus Fragen zur semantischen Plausibilität und Komplexität der SQL-Abfrage und einer vergleichenden Qualitätsanalyse der generierten natürlichsprachlichen Beschreibung der SQL-Abfrage, im Vergleich zu einer Baseline.

Diskussion

Das System erzeugt SQL-Abfragen, die den Anforderungen entsprechen. Die NLG-Pipeline schneidet im Vergleich zur Baseline besser ab. Die Bewerter stellten dennoch eine große Anzahl von qualitativen Fehlern fest. Zukünftige Forschung zur Verbesserung der Textgenerierung ist daher zu empfehlen.

