

SZAKDOLGOZAT KIVONATA

Oldalrangsorolási eljárások és numerikus módszereik

Szász Tünde, matematika BSc

Témavezető: Horváth Róbert

A szakdolgozat témája a különböző oldalrangsorolási eljárások ismertetése és vizsgálata, leginkább az azok által használt numerikus módszerekre koncentrálván. Az oldalrangsorolási eljárásokat a keresőrendszerek (pl. Google) használják, hogy a felhasználók számára áttekinthetőbbé és könnyebben kezelhetővé tegyék a keresési kifejezésre adott találatokat.

Az első részben a PageRank algoritmról van szó, amelyet a Google keresőrendszer használ. Ez a web linkstruktúráját mint Markov-láncot tekinti, és annak stacionárius eloszlását keresi a hatványmódszer segítségével. Ahhoz, hogy a hatványmódszer egyértelműen konvergáljon a megoldáshoz, a linkmátrixot át kell alakítani, hogy primitív, irreducibilis sztochasztikus mátrix legyen. A problémát át lehet alakítani úgy is, hogy egy lineáris egyenletrendszert kelljen megoldani, ekkor különböző iterációs módszereket lehet alkalmazni, mint például a Jacobi- illetve Gauss–Seidel-iteráció.

Ezután következik a két kevésbé elterjedt algoritmus, a HITS és SALSA leírása. Mindkettő két rangsort állít elő. A HITS a keresési kifejezéshez állít elő egy szomszédsági mátrixot és azzal állítja elő a rangsorokat, így egy kisebb mátrixra hajtja végre a hatványmódszert. A SALSA páros gráfot alkot a keresési kifejezésre kapott oldalakból illetve szomszédjaikból és az ebből alakított két Markov-lánc stacionárius eloszlását keresi.

Majd következnek a példák és kísérletek, amik MATLAB segítségével lettek elvégezve. Szemléletes példákon kiderül, hogy a PageRank különböző paraméterválasztás esetén hogyan viselkedik, miképpen befolyásolja a keletkező rangsorvektor pontosságát és a konvergencia gyorsaságát.

Végül pedig egy rövid áttekintés található a Markov-láncokról, illetve a numerikus módszerekről, amelyeket az algoritmusok felhasználnak.