

A Indexação Semântica Latente (ISL) é uma técnica para Recuperação de Informações que objetiva recuperar os documentos com base em seu conteúdo semântico. Considerando-se que o mesmo conceito pode ser expresso usando palavras diferentes (sinonímia) e que uma mesma palavra pode ter vários significados (polissemia), palavras individuais não fornecem uma evidência confiável sobre o significado do texto.

O objetivo do trabalho é estimar o número de dimensões ideal para a Decomposição em Valores Singulares (DVS), utilizada na ISL, a partir de experimentos, utilizando coleções de teste compostas por documentos, consultas e julgamentos de relevância. Esperamos, no fim, ter um método automatizado para estimar esse valor e evitar a utilização de mais dimensões do que o necessário no cálculo da DVS, considerado o seu elevado custo computacional.

Os experimentos realizados geraram os seguintes resultados:

Coleção	Time	Cranfield	CISI	LA Times
Documentos	425	1398	1460	113005
Termos	7600	2629	3121	96537
Maior precisão média	0,3552 (120 dimensões)	0,1720 (170 dimensões)	0,3918 (980 dimensões)	0,2462 (1290 dimensões)

A precisão média representa o quanto a ordenação dos documentos por relevância para cada consulta se assemelha aos julgamentos de relevância.