

第7回小テストの課題

以下の3つのテキストを検索対象にしたブーリアン検索を考える。

テキスト1 Transfer of data blocked by a gate

テキスト2 Debug of soft conducted by a soft engineer

テキスト3 Transfer of data conducted by a engineer

課題1

3つのテキストに対する索引語－文書行列を求めなさい。解答する際には各行、各列が何を表すかを明確にしなさい。また、索引語は各テキストに含まれている単語を全て用いること。単語はスペースで区切られているものとする。ステミングは不要である。

課題2

検索質問「Transfer \wedge engineer」に対するブーリアン検索の過程と検索結果を示しなさい。

課題1の解答

	テキスト1	テキスト2	テキスト3
a	1	1	1
blocked	1	0	0
by	1	1	1
conducted	0	1	1
data	1	0	1
Debug	0	1	0
engineer	0	1	1
gate	1	0	0
of	1	1	1
soft	0	1	0
Transfer	1	0	1

課題2の解答

Transfer の文書ベクトル 1 0 1

engineer の文書ベクトル 0 1 1

101 \wedge

011

=001

よって、テキスト3が検索結果