

Поиск ассоциативных правил в суперкомпьютерных рейтингах Top500 и Top50

П.И. Шумилин, М.Л. Цымблер
Южно-Уральский государственный университет (Челябинск)

Ассоциативные правила

№ п/п	Набор
1	молоко, сахар, хлеб
2	мука, печенье, хлеб
3	молоко, яблоки
4	картофель, молоко, хлеб
5	молоко, сахар, сыр, торт

Поддержка
 $sup(A \rightarrow B) = P(A \cup B)$ – доля транзакций, содержащих наборы A и B совместно.

Достоверность
 $conf(A \rightarrow B) = P(B|A)$ – доля транзакций, содержащих B среди тех, которые содержат A .

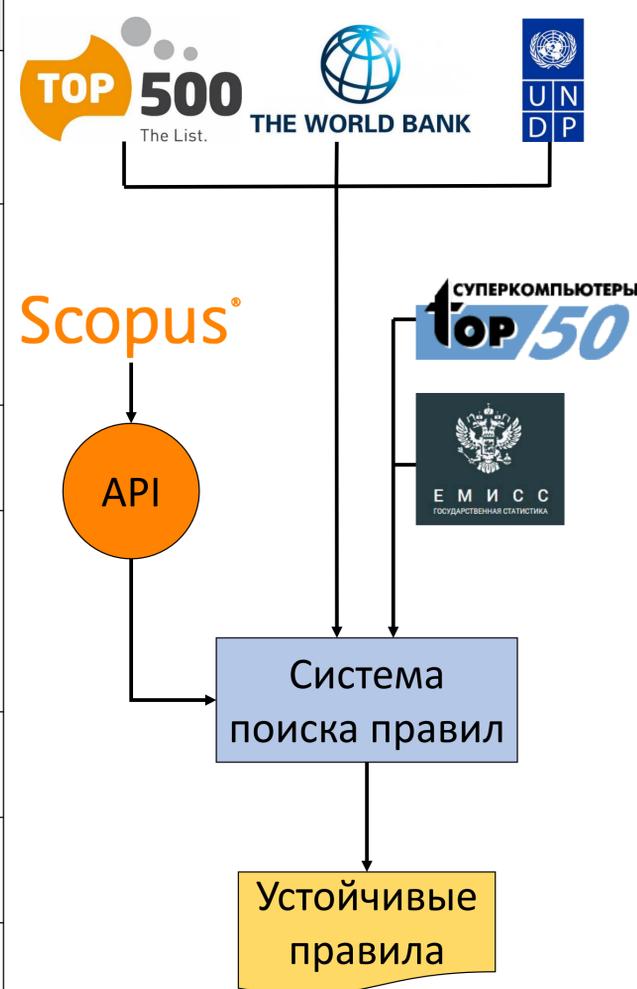
Устойчивое правило
 $sup(A \rightarrow B) \geq minsup$ и $conf(A \rightarrow B) \geq minconf$

$minsup$	$minconf$
0.3	0.6

№ п/п	Правила	sup	$conf$
1	молоко \rightarrow хлеб	2/5	2/4
2	хлеб \rightarrow молоко	2/5	2/3
3	хлеб \rightarrow сахар	1/5	1/3

Подготовка данных

	Top500	Top50
Атрибуты	Количество суперкомпьютеров страны в редакции (Num of Systems)	Производительность суперкомпьютера в научной организации (RMax)
	Валовой внутренний продукт на душу населения (GDPpC), индекс человеческого развития (HDI) страны	Валовой региональный продукт субъекта РФ, где расположена организация
	Количество публикаций по Q1-Q4 Scopus в стране	Количество публикаций по Q1-Q4 Scopus в организации
Источники	Top500, The World Bank, Human Development Reports (UNDP), Scopus	Top50, Единая межведомственная информационно-статистическая система (ЕМИСС), Scopus
Квантификация атрибутов	Децильная	Квартильная
Количество редакций	26	15
Количество транзакций	743	248



Результаты

№ п/п	Рейтинг	Устойчивые правила	
		ЕСЛИ	ТО
1	Top500	[HDI: 7, Q1: 2]	Num of Systems: 2
2		[Num of Systems: 4, Q1: 3, Q2: 3]	GDPpC: 7
3		[HDI: 2, Q4: 1]	Num of Systems: 1
4		[Num of Systems: 2, GDPpC: 5]	Q1: 2
5		[Num of Systems: 9, GDPpC: 10]	Q1: 10
6	Top50	[RMax: 1, Q1: 1]	GRPpC: 1
7		[RMax: 4, GRPpC: 1, Q2: 3]	Q1: 3
8		[RMax: 1, Q1: 3]	GRPpC: 4
9		[GRPpC: 3, Q2: 1, Q3: 1]	RMax: 3
10		[GRPpC: 4, Q1: 2, Q3: 2]	RMax: 2

1. **ЕСЛИ**
низкий уровень ИЧР **И**
большое количество публикаций в Q1
ТО
большое число суперкомпьютеров в рейтинге Top500

8. **ЕСЛИ**
высокая производительность суперкомпьютера **И**
количество публикаций в Q1 ниже среднего уровня
ТО
низкий уровень ВРП на душу населения