

Contents

1	Introduction	1
1.1	The Mind's Similarity Engine	3
1.1.1	Metric Properties	5
1.1.2	Multi-modal Similarity Search	8
1.1.3	Similarity and Grouping: The Diagnosticity Principle	10
1.2	Summary	10
	References	11
2	Fundamentals of Similarity Search	13
2.1	Object Representation	13
2.2	Attribute Types	13
2.3	Comparing Attribute Values: Similarity and Distance	15
2.4	Similarity Measures	16
2.5	Summary	18
	References	18
3	Common Similarity Search Operators	19
3.1	The Similarity Search Framework	19
3.2	Aggregation Functions	21
3.2.1	Weighted Sum and Euclidean Distance	22
3.2.2	Max	23
3.2.3	Min	24
3.2.4	N-Match	24
3.3	Filter Functions	25
3.3.1	Threshold Filter	25
3.3.2	Top-k Filter	26
3.3.3	Skyline Filter	26
3.4	Common Similarity Operators	28
3.4.1	Threshold based Operators	28
3.4.2	Top-k based Operators	29
3.4.3	Skyline Operators	30

3.5	Summary	30
	References	31
4	Categorizing Operators	33
4.1	Types of Similarity Operators	33
4.1.1	Ordered vs. Unordered Result Set	34
4.1.2	All vs. Some: Role of Attributes in Result Set Determination	35
4.1.3	Example Operators under the 4-type Classification	38
4.2	Features: Add-on Functionalities for Similarity Operators	39
4.2.1	Indirection: Forming "Reverse" Operators	39
4.2.2	Chromaticity: Channelizing the Search to Specific Types ...	41
4.2.3	Visibility: Is there a line of sight?	43
4.2.4	Multiple Queries	45
4.2.5	Subspaces: When Some Attributes are Enough	47
4.2.6	Diversity: To avoid Monotony	48
4.2.7	Summary of Features	50
4.3	Summary	51
	References	52
5	Advanced Operators for Similarity Search	55
5.1	Weighted Sum-based Operators	55
5.1.1	Reverse k Nearest Neighbor (RkNN): Indirection on Top- k ..	57
5.1.2	Bichromatic Reverse k Nearest Neighbor (BRkNN)	58
5.1.3	Reverse k Farthest Neighbor (RkFN): Farthest instead of Nearest	60
5.1.4	Constrained Nearest Neighbors: Results must satisfy Constraints	61
5.1.5	Visible kNN and RkNN: Visibility Constraints on Top- k and RkNN	63
5.1.6	Subspace Top- k and Range Queries	65
5.1.7	K Nearest Diverse Neighbors: Diversity-conscious Top- k ...	67
5.1.8	Multi-Query Top- k	68
5.1.9	Summary	70
5.2	Skyline-based Operators	70
5.2.1	Reverse Skyline: Indirection on the Skyline	72
5.2.2	Thick Skyline: Skyline Objects and their Neighbors	73
5.2.3	Constrained Skyline	74
5.2.4	Dynamic Skyline: Skyline in a Transformed Space	75
5.2.5	Skyband: At most k dominators	76
5.2.6	Reverse Skyband: Indirection on the Skyband	77
5.2.7	K -Most Representative Skyline: Maximizing combined dominating Power	78
5.2.8	Top- k Frequent Skylines: Counting Subspaces	79
5.2.9	Skyline k -Groups: Group-level Skyline Evaluation	81
5.2.10	Summary	82

5.3	Other Operators	82
5.3.1	Reverse K -Ranks and Reverse K -Scores	82
5.3.2	Spatial Skyline	84
5.3.3	Hypermatching	85
5.3.4	Summary	87
5.4	Summary	88
	References	88
6	Indexing for Similarity Search Operators	91
6.1	Euclidean Space Indexes	91
6.1.1	Top- k using k -d Trees	92
6.2	Metric Space Indexes	93
6.2.1	Top- k using VP-Trees	94
6.3	Non-metric Space Indexes	96
6.3.1	Middleware Algorithms for Top- k	97
6.3.2	Top- k using AL Trees	99
6.4	Summary	101
	References	102
7	The Road Ahead	105
7.1	Multi-modal Similarity Search	107
7.2	Interoperability between Specialized Indexes	108
7.3	Explaining Similarity Search Results	108
7.4	Leveraging Newer Platforms for Similarity Search	108
7.5	Complex Operator Combinations	109
7.6	Novel Similarity Search Operators	110
7.6.1	N -Match-BB: Any N conditions	110
7.6.2	Multi-Query BRkNN: Multi-Query Extension of BRkNN ..	111
7.6.3	Summary	112
7.7	Summary	112
	References	113
	Index	115