

Analysis of Edge Schema Contained in
jdbc:microsoft:sqlserver://db-
ms.doc.ic.ac.uk;databaseName=dblp

XMLtoEdgeSchema program, Peter McBrien
pjm@doc.ic.ac.uk
<http://www.doc.ic.ac.uk/~pjm>

May 16, 2003

1 Siblings

parent	child	sibling	count
article	author	author	114,171
		pages	2,311
		title	111,072
cdrom		author	9
		cite	206
		ee	269
		journal	2
		number	1
		pages	3
		title	26
		url	282
		volume	2
		year	2,973
cite		cite	45,511
		ee	6
		url	2
editor		title	1
ee		author	2,288
		cdrom	476
		cite	1,425
		ee	23
		number	2
		pages	1
		title	494
		url	19,166
		year	1,373
journal		cdrom	2
		ee	335
		month	2,448
		number	105,750
		url	4,887
		volume	463
		year	609
month		cdrom	1
		ee	1
		volume	1
		year	2,471
number		cdrom	182
		ee	18,847
		journal	189
		month	1
		pages	17
		url	89,448
		volume	308
		year	247
pages		cdrom	3,010
		ee	1,373
		journal	2,462
		url	333

parent	child	sibling	count
article	pages	year	105,264
	publisher	journal	1
		year	3
	title	author	2,311
		cdrom	52
		ee	2
		journal	1,059
		pages	109,699
		publisher	1
		year	1,372
	url	cdrom	517
		cite	187
		ee	4,108
		pages	54
		volume	15
	volume	cdrom	2
		journal	110,762
		month	24
		number	3,075
		pages	5
		publisher	3
		url	311
		year	184
	year	cdrom	19
		ee	14
		journal	21
		month	1
		number	415
		pages	395
		url	35
book	author	author	389
		title	797
	cdrom	cite	3
		url	1
	cite	cite	2,637
		url	1
	editor	editor	79
		title	68
	isbn	cdrom	1
		ee	5
		isbn	6
		publisher	2
		series	33
		url	115
		year	9
	month	isbn	1
	publisher	isbn	11
		publisher	1

parent	child	sibling	count
book	publisher	series	6
		url	1
		year	847
	series	publisher	2
		url	2
		volume	407
	title	publisher	597
		series	261
		url	2
		year	7
	url	cdrom	2
		cite	3
		isbn	2
		publisher	2
		year	1
	volume	isbn	137
		publisher	259
		year	3
	year	cdrom	1
		isbn	647
		month	1
		publisher	3
		series	140
		url	10
dblp	article	article	114,491
		inproceedings	2
		phdthesis	1
		www	1
	book	article	1
		book	819
		incollection	47
	incollection	book	47
		incollection	973
	inproceedings	article	2
		book	1
		inproceedings	220,969
		proceedings	1,074
	mastersthesis	article	1
		mastersthesis	1
		phdthesis	2
		www	1
	phdthesis	mastersthesis	3
		phdthesis	69
	proceedings	inproceedings	1,075
		proceedings	2,109
	www	article	1
		proceedings	1
		www	36
incollection	author	author	1,006

parent	child	sibling	count
incollection	author	title	1,010
	booktitle	crossref	7
		publisher	91
		url	919
		year	3
	chapter	url	2
	cite	cite	60
	crossref	booktitle	32
		cdrom	14
		pages	1
		url	6
	ee	author	38
		cdrom	10
		ee	14
		title	6
		year	20
	isbn	pages	2
		url	26
	pages	booktitle	59
		chapter	2
		crossref	3
		ee	20
		url	38
		year	880
	publisher	year	91
	title	booktitle	41
		pages	960
		year	19
	url	cdrom	29
		cite	5
		crossref	14
		ee	29
		year	7
	year	booktitle	888
		crossref	29
		isbn	28
		pages	39
		url	26
inproceedings	author	author	292,977
		pages	2,775
		title	219,007
	booktitle	cdrom	160
		crossref	109
		ee	69,970
		number	53
		pages	98
		url	151,323
		year	331
	cdrom	booktitle	45

parent	child	sibling	count
inproceedings	cdrom	cdrom	3
		cite	2,594
		crossref	253
		ee	1,234
		pages	118
		url	86
		year	4,958
	cite	cdrom	8
		cite	114,450
		ee	2
	crossref	author	3,118
		booktitle	108,432
		cdrom	651
		cite	4
		crossref	8
		ee	3,152
		pages	3
		title	1
		url	12
		year	3,121
	ee	author	2,757
		booktitle	2
		cdrom	2,919
		cite	3,418
		crossref	115
		ee	120
		pages	9
		title	53
		url	70,170
		year	10,334
	month	year	1
	note	booktitle	3
		cdrom	1
		crossref	1
		url	4
		year	7
	number	ee	25
		url	28
	pages	booktitle	19
		cdrom	5,085
		cite	13
		crossref	3,258
		ee	10,472
		note	12
		url	93
		year	198,712
	title	author	2,775
		cdrom	147
		crossref	89

parent	child	sibling	count
inproceedings	title	month	1
		pages	214,453
		year	4,581
	url	cdrom	907
		cite	343
		crossref	3,564
		ee	3,947
		note	1
		pages	374
		year	1
	year	booktitle	113,545
		cdrom	41
		crossref	108,115
		ee	2
		note	3
		pages	12
		url	328
mastersthesis	author	title	5
		year	5
	year	school	5
phdthesis	author	title	72
		month	3
		number	2
	publisher	isbn	1
		number	1
		school	1
	school	ee	1
		publisher	2
	series	number	1
	title	year	72
	year	month	3
		number	1
		publisher	1
		school	66
		series	1
proceedings	address	publisher	1
		year	3
	author	author	6
		title	1
	booktitle	isbn	3
		month	1
		publisher	287
		series	2,416
		url	11
		volume	7
		year	109
	cite	cite	211
	crossref	isbn	2
	editor	editor	3,661

parent	child	sibling	count
proceedings	editor	title	2,894
	ee	url	3
	isbn	booktitle	10
		isbn	1
		publisher	12
		series	1
		url	2,789
		year	9
	journal	volume	5
	month	year	1
	note	isbn	1
	number	publisher	1
		url	9
		year	5
	publisher	address	3
		booktitle	19
		isbn	24
		number	1
		series	2
		url	13
		volume	6
		year	3,015
	series	booktitle	5
		isbn	3
		publisher	49
		url	1
		volume	2,426
		year	10
	title	booktitle	2,796
		isbn	10
		journal	4
		note	1
		publisher	283
		series	68
		title	1
		url	2
		year	20
	url	booktitle	1
		cite	1
		isbn	50
		journal	1
		publisher	1
		series	1
		year	6
	volume	ee	3
		isbn	5
		number	6
		publisher	2,424
		year	6

parent	child	sibling	count
proceedings	year	address	1
		booktitle	5
		crossref	2
		isbn	2,780
		note	1
		number	9
		publisher	92
		series	6
		url	274
title	i	i	3
		sub	38
		sup	49
	sub	i	17
		sup	40
	sup	i	9
		sub	50
www	author	author	7
		title	7
	booktitle	year	1
	editor	editor	16
		title	5
	title	booktitle	1
		url	27
		year	10
	year	url	11

2 Count of Each Element and Attribute

node	nodetype	count
null	cdata	3,117,602
address	element	4
article	element	114,496
author	element	745,613
book	element	867
booktitle	element	225,903
cdrom	element	14,241
chapter	element	2
cite	element	171,071
crossref	element	118,575
dblp	element	1
editor	element	6,724
ee	element	119,755
href	attribute	3,026
i	element	1,393
incollection	element	1,020
inproceedings	element	222,046
isbn	element	3,712
journal	element	114,501
key	attribute	341,728
label	attribute	72,423
mastersthesis	element	5
month	element	2,480
note	element	18
number	element	109,315
pages	element	331,329
phdthesis	element	72
proceedings	element	3,184
publisher	element	4,114
rating	attribute	61
reviewid	attribute	63
school	element	77
series	element	2,935
sub	element	721
sup	element	1,051
title	element	341,729
tt	element	6
url	element	340,797
volume	element	117,232
www	element	38
year	element	341,701

3 Complex Element Child Analysis

element	cdata	attribute	element
address	4	0	0
article	0	114,620	1,103,479
author	745,613	0	0
book	0	867	8,369
booktitle	225,903	0	0
cdrom	14,241	0	0
chapter	2	0	0
cite	171,071	72,423	0
crossref	118,575	0	0
dblp	0	0	341,728
editor	6,724	0	0
ee	119,755	0	0
i	1,409	0	81
incollection	0	1,020	7,494
inproceedings	0	222,046	1,961,962
isbn	3,712	0	0
journal	114,501	0	0
mastersthesis	0	5	20
month	2,480	0	0
note	18	0	0
number	109,315	0	0
pages	331,329	0	0
phdthesis	0	72	300
proceedings	0	3,184	30,081
publisher	4,114	202	0
school	77	0	0
series	2,935	2,824	0
sub	711	0	22
sup	1,021	0	51
title	344,356	0	3,017
tt	6	0	0
url	340,797	0	0
volume	117,232	0	0
www	0	38	123
year	341,701	0	0

4 Child Cardinality and Data Type

parent	no	child	no	nodetype	datatype	size	min	max	avg
address	4	null	4	cdata	string	4	1	1	1.000
article	114,496	author	227,554	element	string	113,383	0	52	1.987
		cdrom	4,262	element	string	4,261	0	2	0.037
		cite	47,329	element	complex	1,818	0	348	0.413
		editor	1	element	string	1	0	1	0.000
		ee	27,794	element	string	26,969	0	3	0.243
		journal	114,496	element	string	114,496	1	1	1.000
		key	114,496	attribute	string	114,496	1	1	1.000
		month	2,474	element	string	2,474	0	1	0.022
		number	109,243	element	string	109,243	0	1	0.954
		pages	112,485	element	string	112,484	0	2	0.982
		publisher	4	element	string	4	0	1	0.000
		rating	61	attribute	string	61	0	1	0.001
		reviewid	63	attribute	string	63	0	1	0.001
		title	114,496	element	complex	114,496	1	1	1.000
		url	114,464	element	string	114,464	0	1	1.000
		volume	114,381	element	string	114,381	0	1	0.999
		year	114,496	element	string	114,496	1	1	1.000
author	745,613	null	745,613	cdata	string	745,613	1	1	1.000
book	867	author	1,186	element	string	797	0	18	1.368
		cdrom	4	element	string	4	0	1	0.005
		cite	2,643	element	complex	6	0	741	3.048
		editor	147	element	string	68	0	6	0.170
		ee	5	element	string	5	0	1	0.006
		isbn	804	element	string	798	0	2	0.927
		key	867	attribute	string	867	1	1	1.000
		month	1	element	string	1	0	1	0.001
		publisher	866	element	complex	865	0	2	0.999
		series	440	element	complex	440	0	1	0.507
		title	867	element	string	867	1	1	1.000
		url	132	element	string	132	0	1	0.152
		volume	407	element	string	407	0	1	0.469
		year	867	element	string	867	1	1	1.000
booktitle	225,903	null	225,903	cdata	string	225,903	1	1	1.000
cdrom	14,241	null	14,241	cdata	string	14,241	1	1	1.000
chapter	2	null	2	cdata	byte	2	1	1	1.000
cite	171,071	null	171,071	cdata	string	171,071	1	1	1.000
		label	72,423	attribute	string	72,423	0	1	0.423
crossref	118,575	null	118,575	cdata	string	118,575	1	1	1.000
dblp	1	article	114,496	element	complex	1	114,496	114,496	114,496.000
		book	867	element	complex	1	867	867	867.000
		incollection	1,020	element	complex	1	1,020	1,020	1,020.000
		inproceedings	222,046	element	complex	1	222,046	222,046	222,046.000
		mastersthesis	5	element	complex	1	5	5	5.000
		phdthesis	72	element	complex	1	72	72	72.000
		proceedings	3,184	element	complex	1	3,184	3,184	3,184.000
		www	38	element	complex	1	38	38	38.000
editor	6,724	null	6,724	cdata	string	6,724	1	1	1.000
ee	119,755	null	119,755	cdata	string	119,755	1	1	1.000

parent	no	child	no	nodetype	datatype	size	min	max	avg	std
i	1,393	null	1,409	cdata	string	1,386	0	5	1.011	0.173
		sub	46	element	string	42	0	3	0.033	0.198
		sup	35	element	string	33	0	2	0.025	0.165
incollection	1,020	author	2,016	element	string	1,010	0	16	1.976	1.360
		booktitle	1,020	element	string	1,020	1	1	1.000	0.000
		cdrom	53	element	string	53	0	1	0.052	0.222
		chapter	2	element	string	2	0	1	0.002	0.044
		cite	65	element	complex	5	0	20	0.064	1.027
		crossref	53	element	string	53	0	1	0.052	0.222
		ee	107	element	string	93	0	2	0.105	0.349
		isbn	28	element	string	28	0	1	0.027	0.163
		key	1,020	attribute	string	1,020	1	1	1.000	0.000
		pages	1,002	element	string	1,002	0	1	0.982	0.132
		publisher	91	element	complex	91	0	1	0.089	0.285
		title	1,020	element	complex	1,020	1	1	1.000	0.000
		url	1,017	element	string	1,017	0	1	0.997	0.054
		year	1,020	element	string	1,020	1	1	1.000	0.000
inproceedings	222,046	author	514,759	element	string	221,782	0	33	2.318	1.372
		booktitle	222,046	element	string	222,046	1	1	1.000	0.000
		cdrom	9,922	element	string	9,918	0	2	0.045	0.207
		cite	120,822	element	complex	6,370	0	137	0.544	3.656
		crossref	118,520	element	string	118,507	0	2	0.534	0.499
		ee	91,845	element	string	90,444	0	2	0.414	0.505
		key	222,046	attribute	string	222,046	1	1	1.000	0.000
		month	1	element	string	1	0	1	0.000	0.002
		note	16	element	string	16	0	1	0.000	0.008
		number	53	element	string	53	0	1	0.000	0.015
		pages	217,842	element	string	217,841	0	2	0.981	0.136
		title	222,046	element	complex	222,046	1	1	1.000	0.000
		url	222,044	element	string	222,042	0	2	1.000	0.005
		year	222,046	element	string	222,046	1	1	1.000	0.000
isbn	3,712	null	3,712	cdata	string	3,712	1	1	1.000	0.000
journal	114,501	null	114,501	cdata	string	114,501	1	1	1.000	0.000
mastersthesis	5	author	5	element	string	5	1	1	1.000	0.000
		key	5	attribute	string	5	1	1	1.000	0.000
		school	5	element	string	5	1	1	1.000	0.000
		title	5	element	string	5	1	1	1.000	0.000
		year	5	element	string	5	1	1	1.000	0.000
month	2,480	null	2,480	cdata	string	2,480	1	1	1.000	0.000
note	18	null	18	cdata	string	18	1	1	1.000	0.000
number	109,315	null	109,315	cdata	string	109,315	1	1	1.000	0.000
pages	331,329	null	331,329	cdata	string	331,329	1	1	1.000	0.000
phdthesis	72	author	72	element	string	72	1	1	1.000	0.000
		ee	1	element	string	1	0	1	0.014	0.118
		isbn	1	element	string	1	0	1	0.014	0.118
		key	72	attribute	string	72	1	1	1.000	0.000
		month	3	element	string	3	0	1	0.042	0.201
		number	3	element	string	3	0	1	0.042	0.201
		publisher	3	element	string	3	0	1	0.042	0.201

parent	no	child	no	nodetype	datatype	size	min	max	avg	std
phdthesis	72	school	72	element	string	72	1	1	1.000	0.000
		series	1	element	string	1	0	1	0.014	0.118
		title	72	element	string	72	1	1	1.000	0.000
		year	72	element	string	72	1	1	1.000	0.000
proceedings	3,184	address	4	element	string	4	0	1	0.001	0.035
		author	7	element	string	1	0	7	0.002	0.124
		booktitle	2,836	element	string	2,836	0	1	0.891	0.312
		cite	212	element	complex	1	0	212	0.067	3.757
		crossref	2	element	string	2	0	1	0.001	0.025
		editor	6,555	element	string	2,894	0	17	2.059	1.314
		ee	3	element	string	3	0	1	0.001	0.031
		isbn	2,879	element	string	2,878	0	2	0.904	0.295
		journal	5	element	string	5	0	1	0.002	0.040
		key	3,184	attribute	string	3,184	1	1	1.000	0.000
		month	1	element	string	1	0	1	0.000	0.018
		note	2	element	string	2	0	1	0.001	0.025
		number	16	element	string	16	0	1	0.005	0.071
		publisher	3,150	element	complex	3,150	0	1	0.989	0.103
		series	2,494	element	complex	2,494	0	1	0.783	0.412
		title	3,185	element	complex	3,184	1	2	1.000	0.018
		url	3,102	element	string	3,102	0	1	0.974	0.158
		volume	2,444	element	string	2,444	0	1	0.768	0.422
		year	3,184	element	string	3,184	1	1	1.000	0.000
publisher	4,114	null	4,114	cdata	string	4,114	1	1	1.000	0.000
		href	202	attribute	string	202	0	1	0.049	0.216
school	77	null	77	cdata	string	77	1	1	1.000	0.000
series	2,935	null	2,935	cdata	string	2,935	1	1	1.000	0.000
		href	2,824	attribute	string	2,824	0	1	0.962	0.191
sub	721	null	711	cdata	string	711	0	1	0.986	0.117
		i	10	element	string	10	0	1	0.014	0.117
		sub	1	element	string	1	0	1	0.001	0.037
		sup	11	element	complex	11	0	1	0.015	0.123
sup	1,051	null	1,021	cdata	string	1,018	0	2	0.971	0.183
		i	39	element	string	39	0	1	0.037	0.189
		sub	9	element	string	9	0	1	0.009	0.092
		sup	3	element	string	3	0	1	0.003	0.053
title	341,729	null	344,356	cdata	string	341,678	0	15	1.008	0.109
		i	1,344	element	complex	1,084	0	15	0.004	0.081
		sub	665	element	complex	558	0	14	0.002	0.060
		sup	1,002	element	complex	902	0	6	0.003	0.061
		tt	6	element	string	6	0	1	0.000	0.004
tt	6	null	6	cdata	string	6	1	1	1.000	0.000
url	340,797	null	340,797	cdata	string	340,797	1	1	1.000	0.000
volume	117,232	null	117,232	cdata	string	117,232	1	1	1.000	0.000
www	38	author	14	element	string	7	0	5	0.368	0.998
		booktitle	1	element	string	1	0	1	0.026	0.162
		editor	21	element	string	5	0	6	0.553	1.519
		key	38	attribute	string	38	1	1	1.000	0.000
		title	38	element	string	38	1	1	1.000	0.000

parent	no	child	no	nodetype	datatype	size	min	max	avg	std
www	38	url	38	element	string	38	1	1	1.000	0.000
		year	11	element	string	11	0	1	0.289	0.460
year	341,701	null	341,700	cdata	short	341,700	0	1	1.000	0.002