

## ▼ 1 Investigating Hybrid Strategies for Systematic Literature Review

### Experiment 1

#### Study

Approaches to strategic alignment of software process improvement: A systematic literature review

Francisco J.S.Vasconcellos, Geraldo B.Landre, José Adson O.G.Cunha, Juliano L.Oliveira, Ronaldo A.Ferreira, Auri M.R.Vincenzid

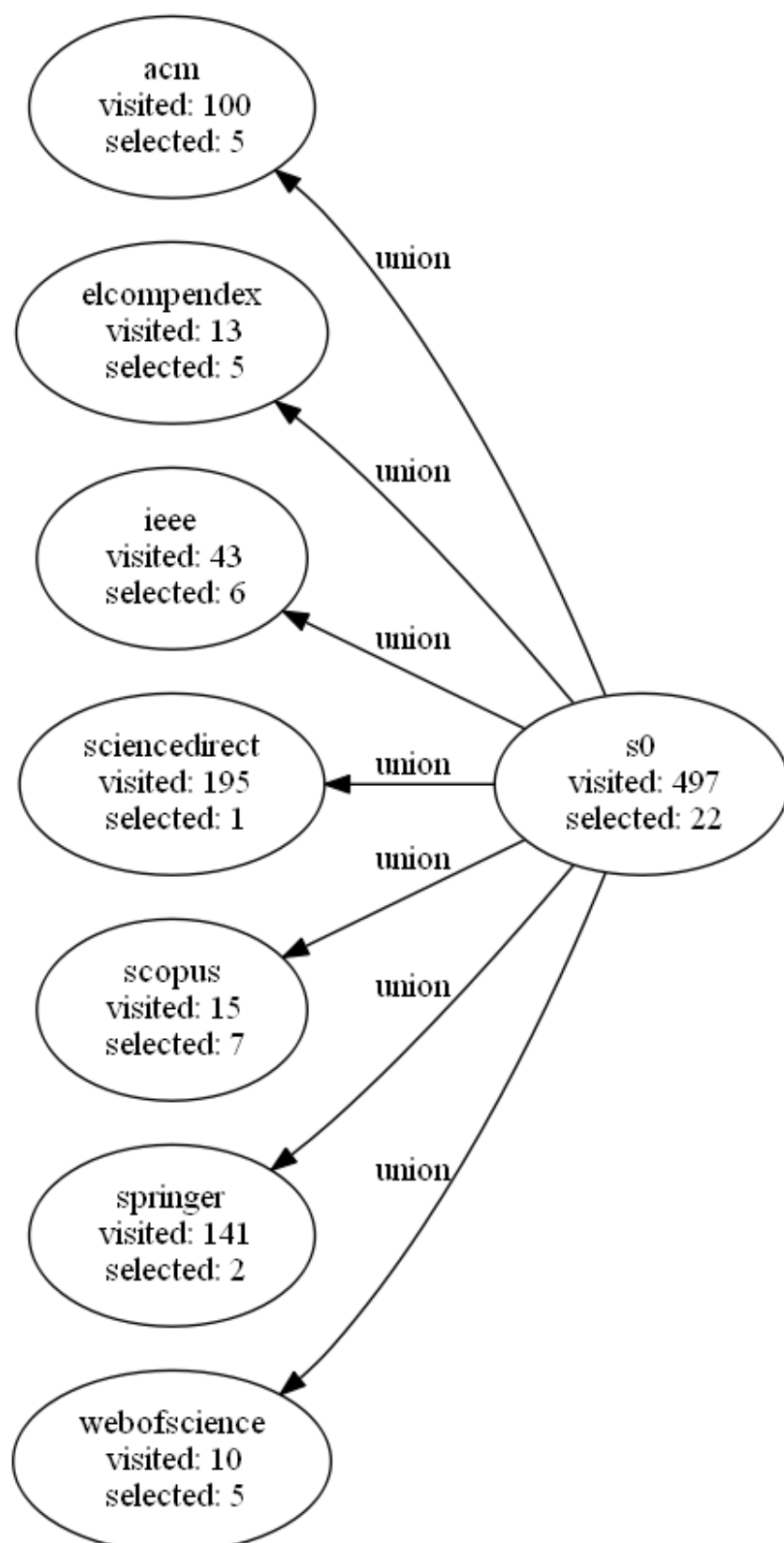
The Journal of Systems and Software, 2017

## ▼ 2 Strategies

### ▼ 2.1 Strategy 1 - Database Search (Search)

Total of selected studies = 22 (Seed Set) + 29 (Snowballing) = 51

Provenance Graph - Strategy 1 - Database Search (Search)



## Report Strategy 1 - Database Search (Search)

Total of selected studies in the SLR = 51

Precision = Selected/ Visited

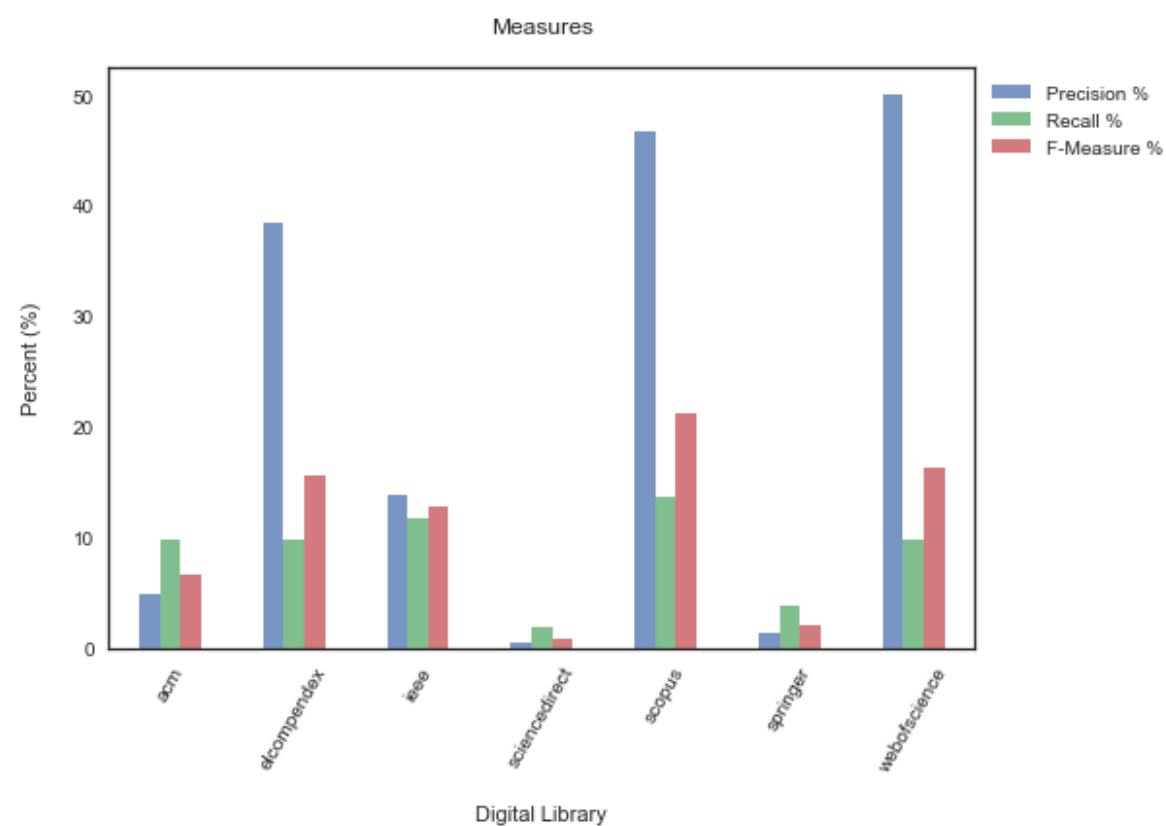
Recall = Selected/ Total of selected studies in the SLR

F-Measure = Harmonic Mean of Precision and Recall

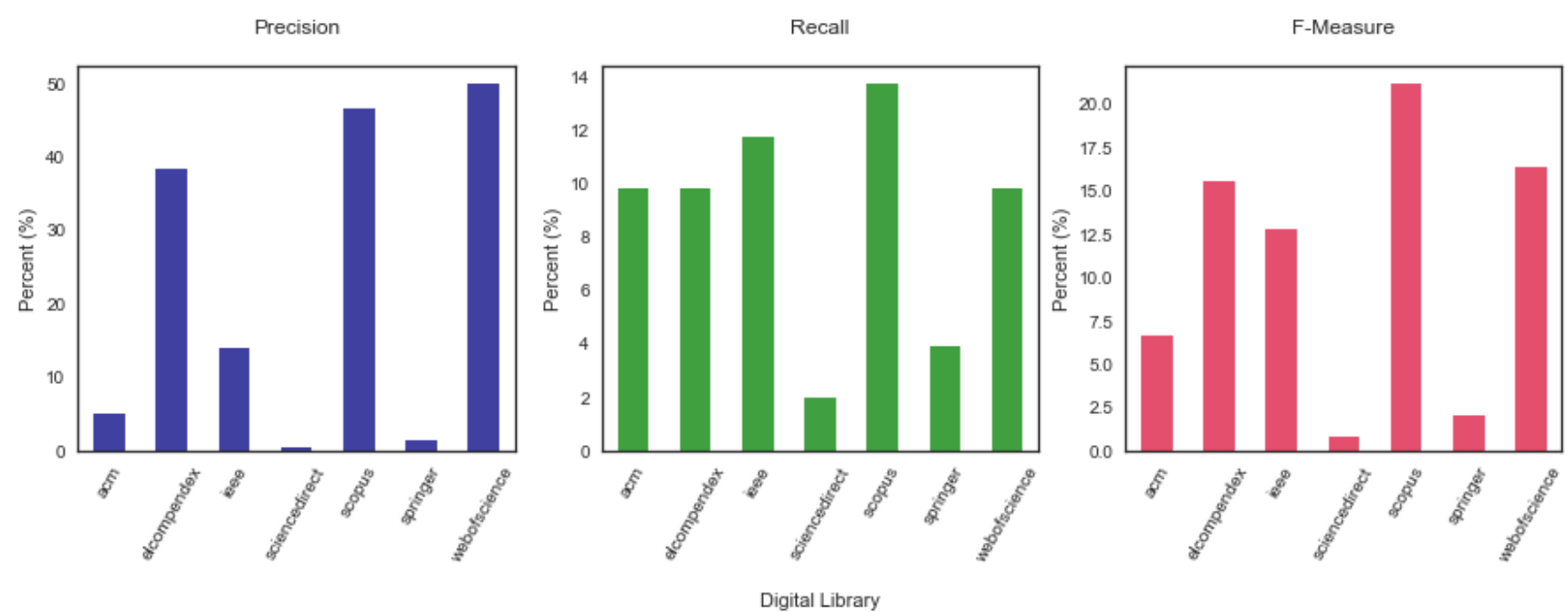
Out[8]:

Digital Library	Visited	Selected	Precision %	Recall %	F-Measure %
acm	100	5	5.00	9.80	6.62
elcompendex	13	5	38.46	9.80	15.62
ieee	43	6	13.95	11.76	12.77
sciencedirect	195	1	0.51	1.96	0.81
scopus	15	7	46.67	13.73	21.21
springer	141	2	1.42	3.92	2.08
webofscience	10	5	50.00	9.80	16.39

Bar Chart Strategy 1



Bar Chart Strategy 1



## Report Strategy 1 Measures (Accumulated)

Out[12]:

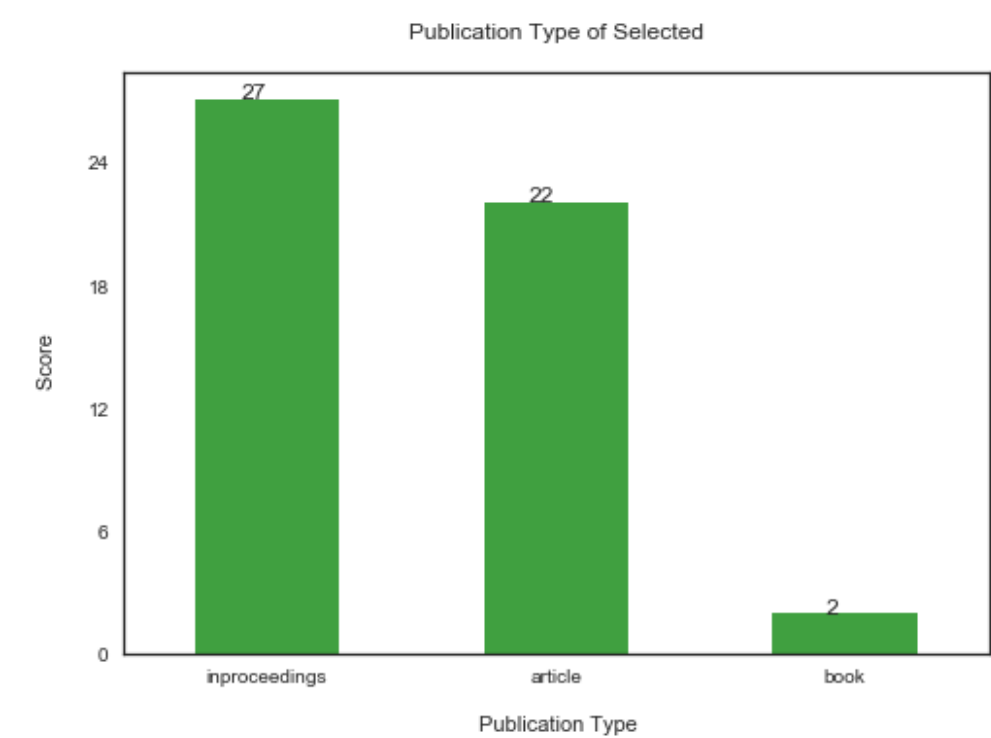
Visited	Selected	Precision %	Recall %	F-Measure %
497	22	4.43	43.14	8.03

Report Strategy 1 - Measures of Strategy - Union of all digital libraries (accumulated)

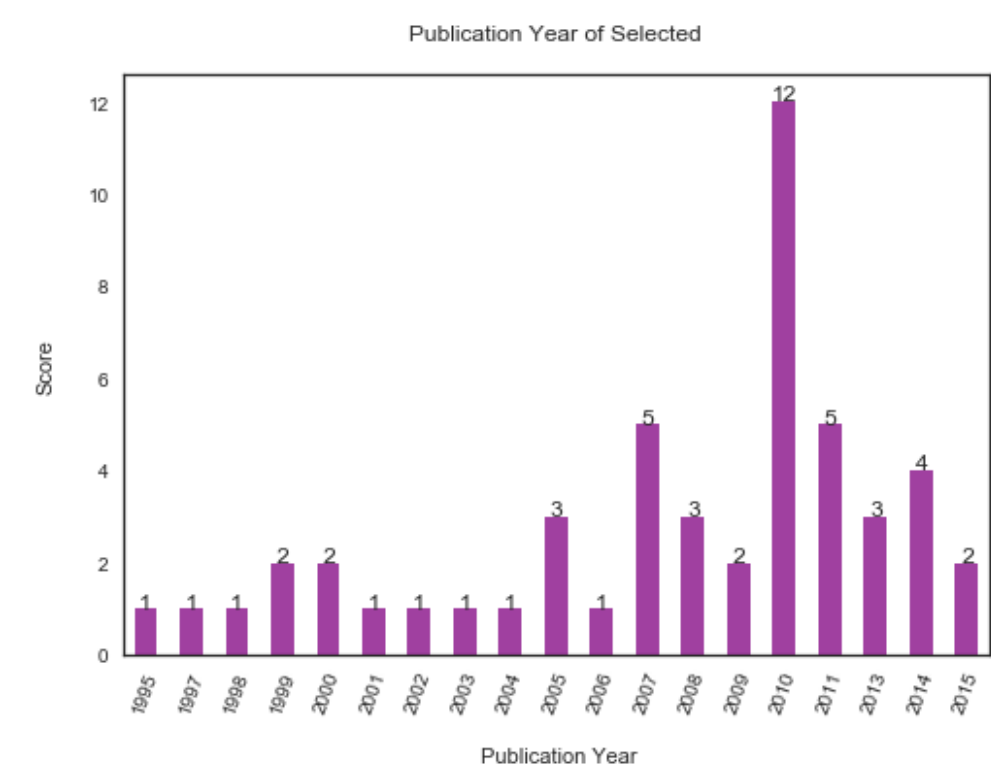
Out[13]:

Strategy	Visited	Selected	Precision %	Recall %	F-Measure %
Search	497	22	4.43	43.14	8.03

Bar Chart Publications Type - Strategy 1

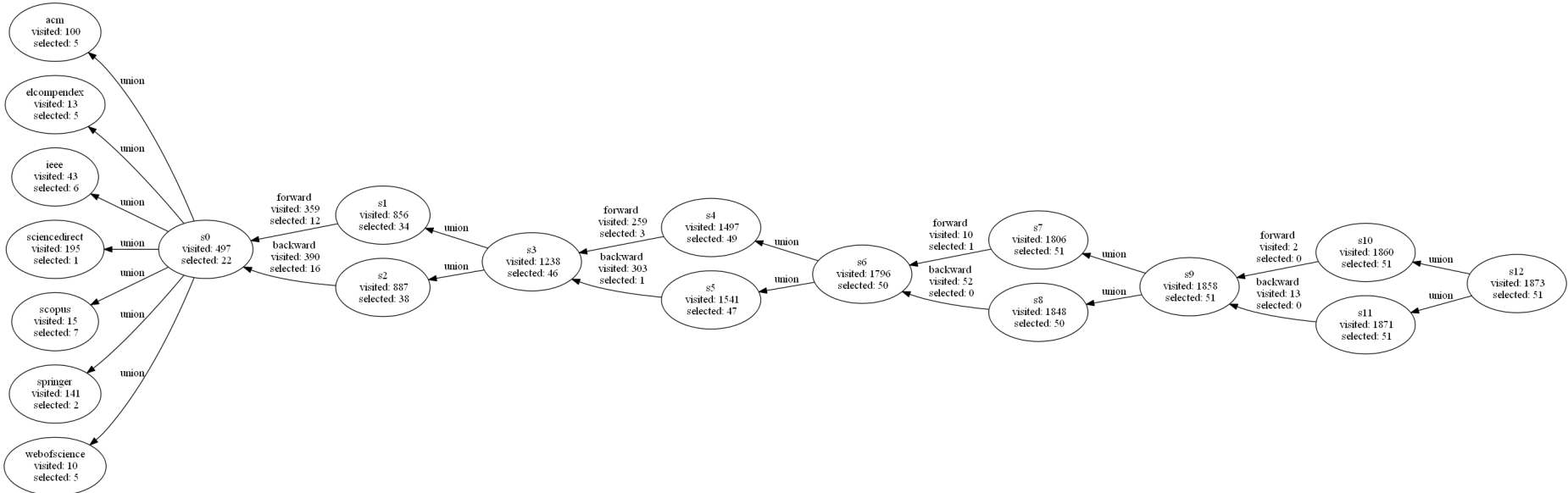


Bar Chart Publications Year - Strategy 1



▼ 2.2 Strategy 2 - Database Search + Snowballing of Wohlin Guideline (Search + SB)

Provenance Graph - Strategy 2 - Database Search + Snowballing of Wohlin Guideline (Search + SB)



Report Strategy 2 - Database Search + Snowballing of Wohlin Guideline (Search + SB)

Total of selected studies in the SLR = 51

Precision = Selected/ Visited  
Recall = Selected/ Total of selected studies in the SLR  
F-Measure = Harmonic Mean of Precision and Recall

Out[18]:

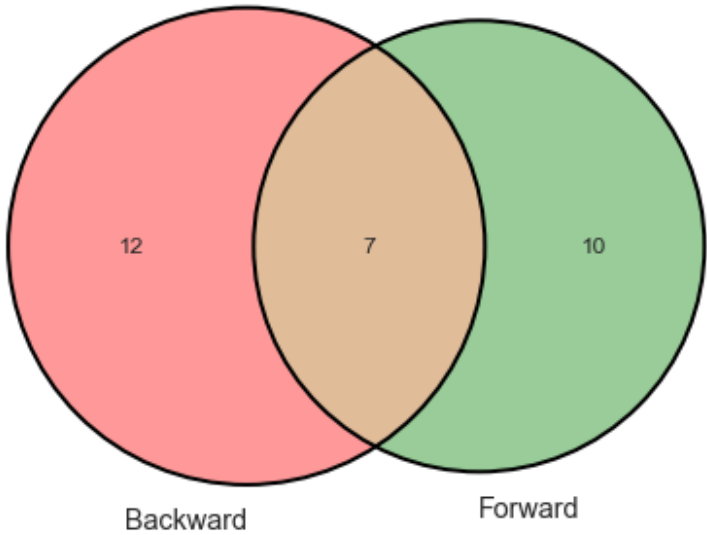
Digital Library/State	Delta Visited	Delta Selected	Delta Precision%	Delta Recall%	Delta F_Measure%	Operation	Accumulated Visited	Accumulated Selected	Accumulated Precision%	Accumulated Recall%	Accumulated F_Measure%
acm	100	5	5.00	9.80	6.62	-	100	5	5.00	9.80	6.62
elcompindex	13	5	38.46	9.80	15.62	-	13	5	38.46	9.80	15.62
ieee	43	6	13.95	11.76	12.77	-	43	6	13.95	11.76	12.77
sciencedirect	195	1	0.51	1.96	0.81	-	195	1	0.51	1.96	0.81
scopus	15	7	46.67	13.73	21.21	-	15	7	46.67	13.73	21.21
springer	141	2	1.42	3.92	2.08	-	141	2	1.42	3.92	2.08
webofscience	10	5	50.00	9.80	16.39	-	10	5	50.00	9.80	16.39
s0	0	0	nan	0.00	0.00	union	497	22	4.43	43.14	8.03
s1	359	12	3.34	23.53	5.85	forward	856	34	3.97	66.67	7.50
s2	390	16	4.10	31.37	7.26	backward	887	38	4.28	74.51	8.10
s3	0	0	nan	0.00	0.00	union	1238	46	3.72	90.20	7.14
s4	259	3	1.16	5.88	1.94	forward	1497	49	3.27	96.08	6.33
s5	303	1	0.33	1.96	0.56	backward	1541	47	3.05	92.16	5.90
s6	0	0	nan	0.00	0.00	union	1796	50	2.78	98.04	5.41
s7	10	1	10.00	1.96	3.28	forward	1806	51	2.82	100.00	5.49
s8	52	0	0.00	0.00	0.00	backward	1848	50	2.71	98.04	5.27
s9	0	0	nan	0.00	0.00	union	1858	51	2.74	100.00	5.34
s10	2	0	0.00	0.00	0.00	forward	1860	51	2.74	100.00	5.34
s11	13	0	0.00	0.00	0.00	backward	1871	51	2.73	100.00	5.31
s12	0	0	nan	0.00	0.00	union	1873	51	2.72	100.00	5.30

Report Strategy 2 - Measures of Strategy (accumulated)

Out[19]:

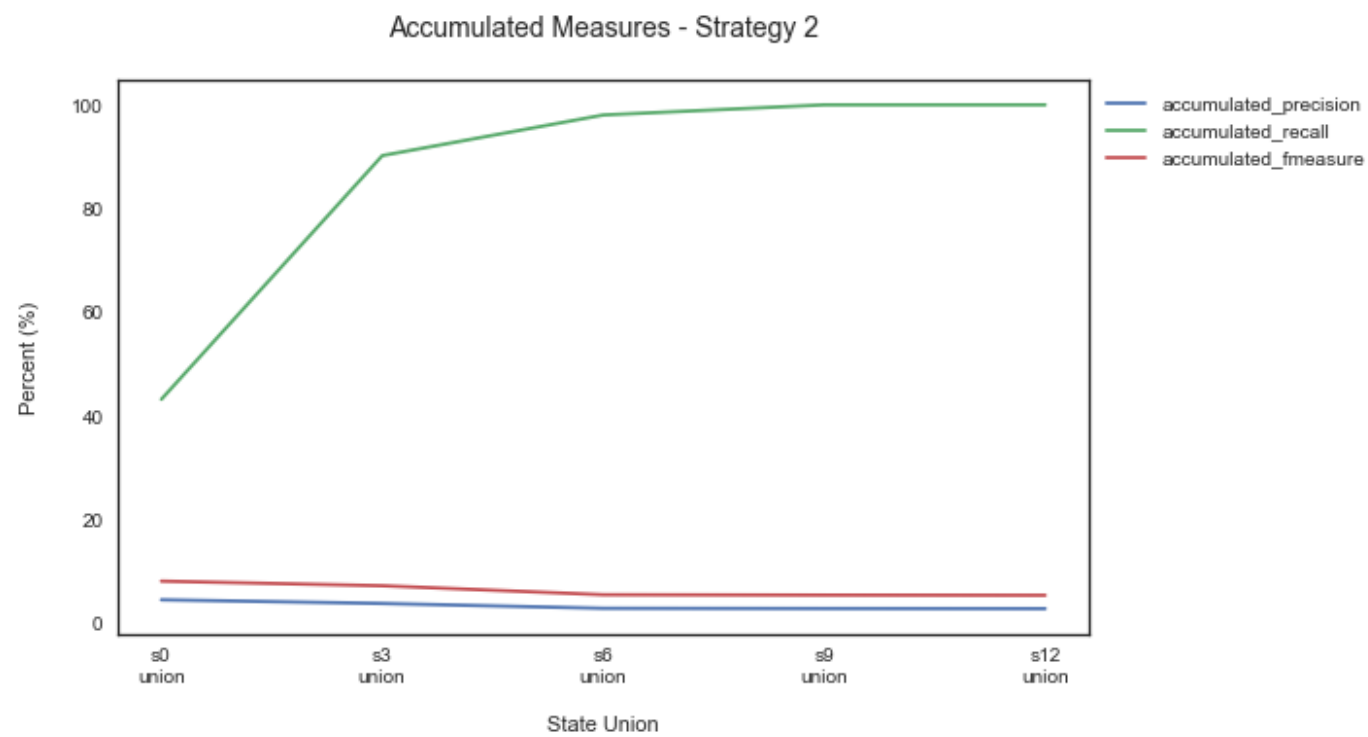
Strategy	Visited	Selected	Precision %	Recall %	F-Measure %
Search + SB	1873	51	2.72	100.00	5.30

Diagrama de Venn - Strategy 2



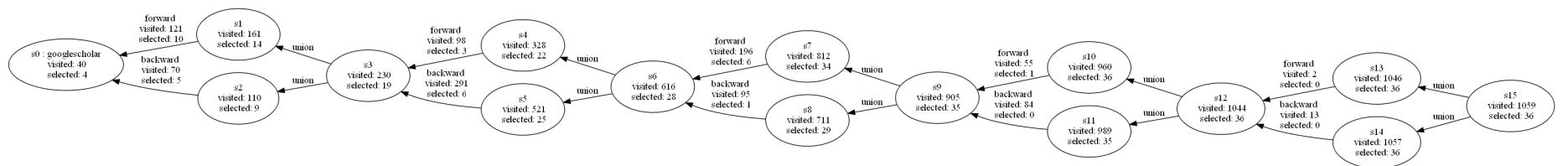
Report Diagrama de Venn

Backward ['hinley1995a', 'basili2007a', 'debou2000a', 'kowalczyk2010a', 'münch2013a', 'liu2005a', 'murugappan2003a', 'basili2007c', 'martins2007b', 'mccoy1998a', 'basili2007b', 'lepmets2014a', 'messnarz1999a', 'trendowicz2011a', 'armbrust2010a', 'mandić2010d', 'münch2013c', 'mcloughlin2010a', 'liu2006a']  
Forward ['sun2010a', 'kowalczyk2010a', 'münch2013a', 'armbrust2010b', 'trendowicz2014a', 'cocozza2014a', 'stallinger2013a', 'unterkalmsteiner2014a', 'karlström2002a', 'birkhölzer2011a', 'trendowicz2011a', 'armbrust2010a', 'trienekens2004a', 'mandić2010d', 'münch2013c', 'mcloughlin2010a', 'mcloughlin2010b']



### 2.3 Strategy 3 - Informal Search (Google Scholar) + Snowballing of Wohlin Guideline = (GS + SB)

Provenance Graph - Strategy 3 - Informal Search (Google Scholar) + Snowballing (Guideline Wohlin)



Report Strategy 3 - Informal Search in Google Scholar + Snowballing of Wohlin Guideline = (GS + SB)

Total of selected studies in the SLR = 51

Precision = Selected/ Visited

Recall = Selected/ Total of selected studies in the SLR

F-Measure = Harmonic Mean of Precision and Recall

Out[25]:

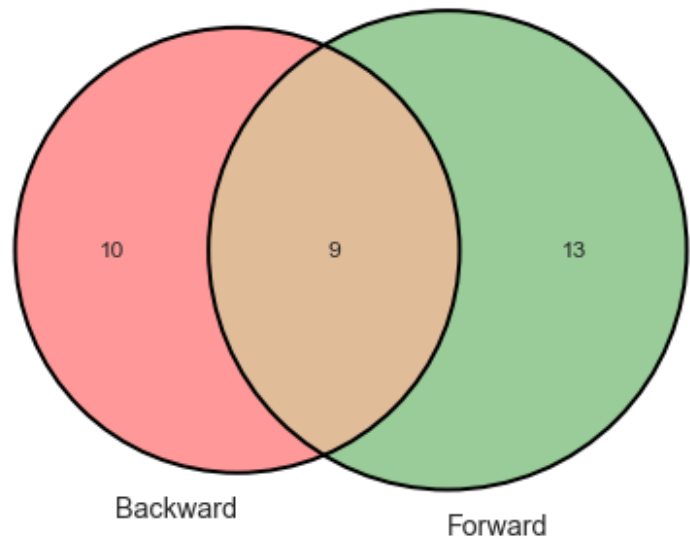
Scholar=s0 /State	Delta Visited	Delta Selected	Delta Precision%	Delta Recall%	Delta F_Measure%	Operation	Accumulated Visited	Accumulated Selected	Accumulated Precision%	Accumulated Recall%	Accumulated F_Measure%
s0 : googlescholar	40	4	10.00	7.84	8.79	-	40	4	10.00	7.84	8.79
s1	121	10	8.26	19.61	11.63	forward	161	14	8.70	27.45	13.21
s2	70	5	7.14	9.80	8.26	backward	110	9	8.18	17.65	11.18
s3	0	0	nan	0.00	0.00	union	230	19	8.26	37.25	13.52
s4	98	3	3.06	5.88	4.03	forward	328	22	6.71	43.14	11.61
s5	291	6	2.06	11.76	3.51	backward	521	25	4.80	49.02	8.74
s6	0	0	nan	0.00	0.00	union	616	28	4.55	54.90	8.40
s7	196	6	3.06	11.76	4.86	forward	812	34	4.19	66.67	7.88
s8	95	1	1.05	1.96	1.37	backward	711	29	4.08	56.86	7.61
s9	0	0	nan	0.00	0.00	union	905	35	3.87	68.63	7.32
s10	55	1	1.82	1.96	1.89	forward	960	36	3.75	70.59	7.12
s11	84	0	0.00	0.00	0.00	backward	989	35	3.54	68.63	6.73
s12	0	0	nan	0.00	0.00	union	1044	36	3.45	70.59	6.58
s13	2	0	0.00	0.00	0.00	forward	1046	36	3.44	70.59	6.56
s14	13	0	0.00	0.00	0.00	backward	1057	36	3.41	70.59	6.50
s15	0	0	nan	0.00	0.00	union	1059	36	3.40	70.59	6.49

Report Strategy 3 - Measures of Strategy (accumulated)

Out[26]:

Strategy	Visited	Selected	Precision %	Recall %	F-Measure %
GS + SB	1059	36	3.40	70.59	6.49

Diagrama de Venn - Strategy 3

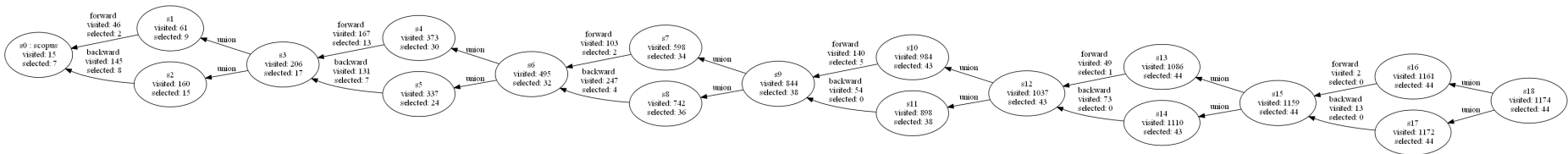


Report Diagrama de Venn

Backward ['basili2007a', 'debou2000a', 'kowalczyk2010a', 'münch2013a', 'mandić2010a', 'liu2005a', 'trienekens2009a', 'murug appan2003a', 'basili2007c', 'mccoy1998a', 'becker2008b', 'basili2007b', 'messnarz1999a', 'trendowicz2011a', 'armbrust2010 a', 'mandić2010d', 'münch2013c', 'mcloughlin2010a', 'liu2006a']  
Forward ['sun2010a', 'kowalczyk2010a', 'münch2013a', 'armbrust2010b', 'trendowicz2014a', 'mandić2010a', 'cocozza2014a', 'pe tersen2015a', 'becker2008a', 'unterkalmsteiner2014a', 'birkhölzer2011a', 'becker2008b', 'wang2005a', 'guzmán2010a', 'trendo wicz2011a', 'armbrust2010a', 'trienekens2004a', 'mandić2010d', 'trienekens2005a', 'mcloughlin2010a', 'münch2013c', 'mclough lin2010b']

2.4 Strategy 4 - Scopus + Snowballing of Wohlin Guideline (Scopus + SB)

Provenance Graph - Strategy 4 - Scopus + Snowballing of Wohlin Guideline (Scopus + SB)



Report Strategy 4 - Scopus + Snowballing of Wohlin Guideline (Scopus + SB)

Total of selected studies in the SLR = 51

Precision = Selected/ Visited  
Recall = Selected/ Total of selected studies in the SLR  
F-Measure = Harmonic Mean of Precision and Recall

Out[31]:

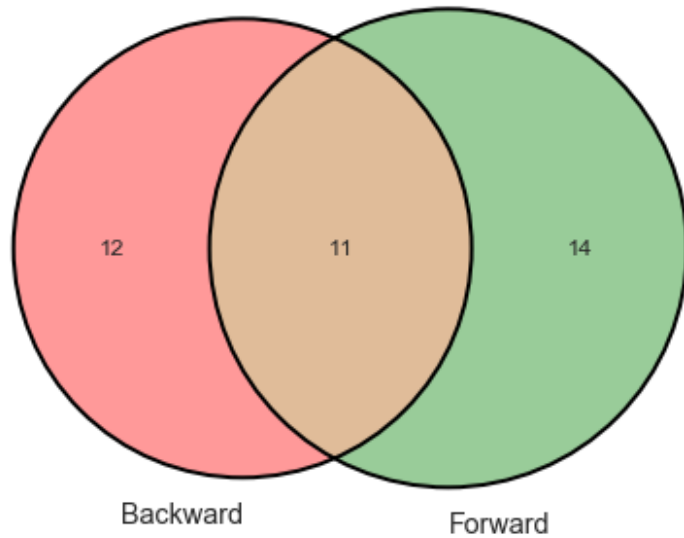
Scopus=s0 /State	Delta Visited	Delta Selected	Delta Precision%	Delta Recall%	Delta F_Measure%	Operation	Accumulated Visited	Accumulated Selected	Accumulated Precision%	Accumulated Recall%	Accumulated F_Measure%
s0 : scopus	15	7	46.67	13.73	21.21	-	15	7	46.67	13.73	21.21
s1	46	2	4.35	3.92	4.12	forward	61	9	14.75	17.65	16.07
s2	145	8	5.52	15.69	8.16	backward	160	15	9.38	29.41	14.22
s3	0	0	nan	0.00	0.00	union	206	17	8.25	33.33	13.23
s4	167	13	7.78	25.49	11.93	forward	373	30	8.04	58.82	14.15
s5	131	7	5.34	13.73	7.69	backward	337	24	7.12	47.06	12.37
s6	0	0	nan	0.00	0.00	union	495	32	6.46	62.75	11.72
s7	103	2	1.94	3.92	2.60	forward	598	34	5.69	66.67	10.48
s8	247	4	1.62	7.84	2.68	backward	742	36	4.85	70.59	9.08
s9	0	0	nan	0.00	0.00	union	844	38	4.50	74.51	8.49
s10	140	5	3.57	9.80	5.24	forward	984	43	4.37	84.31	8.31
s11	54	0	0.00	0.00	0.00	backward	898	38	4.23	74.51	8.01
s12	0	0	nan	0.00	0.00	union	1037	43	4.15	84.31	7.90
s13	49	1	2.04	1.96	2.00	forward	1086	44	4.05	86.27	7.74
s14	73	0	0.00	0.00	0.00	backward	1110	43	3.87	84.31	7.41
s15	0	0	nan	0.00	0.00	union	1159	44	3.80	86.27	7.27
s16	2	0	0.00	0.00	0.00	forward	1161	44	3.79	86.27	7.26
s17	13	0	0.00	0.00	0.00	backward	1172	44	3.75	86.27	7.20
s18	0	0	nan	0.00	0.00	union	1174	44	3.75	86.27	7.18

#### Report Strategy 4 - Measures of Strategy (accumulated)

Out[32]:

Strategy	Visited	Selected	Precision %	Recall %	F-Measure %
Scopus + SB	1174	44	3.75	86.27	7.18

Diagrama de Venn - Strategy 4

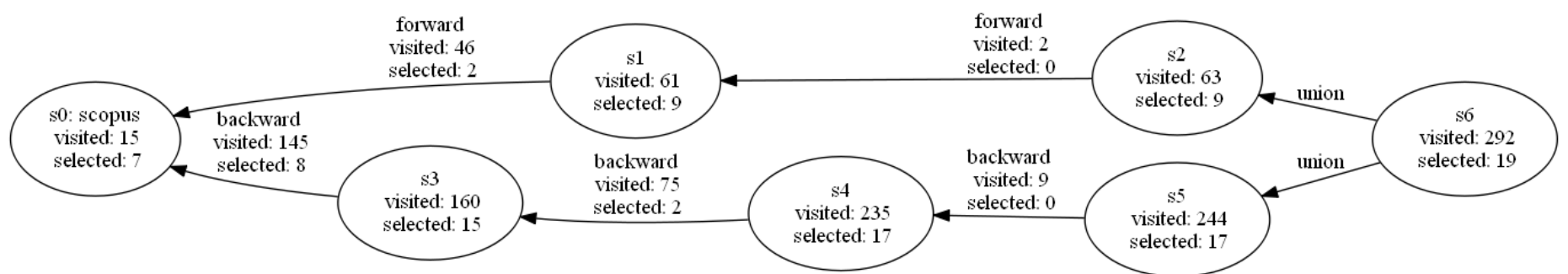


#### Report Diagrama de Venn

Backward ['hinley1995a', 'basili2007a', 'martins2007b', 'basili2010a', 'debou2000a', 'mccoy1998a', 'kowalczyk2010a', 'becke  
r2008b', 'mandić2010b', 'münch2013a', 'basili2007b', 'mandić2010a', 'liu2005a', 'messnarz1999a', 'trendowicz2011a', 'triene  
kens2009a', 'armbrust2010a', 'mandić2010d', 'münch2013c', 'mcloughlin2010a', 'liu2006a', 'murugappan2003a', 'basili2007c']  
Forward ['sun2010a', 'kowalczyk2010a', 'münch2013a', 'armbrust2010b', 'trendowicz2014a', 'mandić2010a', 'cocozza2014a', 'st  
allinger2013a', 'petersen2015a', 'unterkalmsteiner2014a', 'barreto2010a', 'basili2010a', 'birkhölzer2011a', 'becker2008b',  
'mandić2010b', 'wang2005a', 'trendowicz2011a', 'armbrust2010a', 'trienekens2004a', 'mandić2010d', 'trienekens2005a', 'mclou  
ughlin2010a', 'münch2013c', 'martins2008a', 'mcloughlin2010b']

## ▼ 2.5 Strategy 5 - Scopus + Parallel Snowballing of Paper ESEM2017 (Scopus + BS // FS)

Provenance Graph - Strategy 5 - Scopus + Parallel Snowballing of Paper ESEM2017 (Scopus + BS // FS)



#### Report Strategy 5 - Scopus + Parallel Snowballing of Paper ESEM2017 (Scopus + BS // FS)

Total of selected studies in the SLR = 51

Precision = Selected/ Visited

Recall = Selected/ Total of selected studies in the SLR

F-Measure = Harmonic Mean of Precision and Recall

Out[37]:

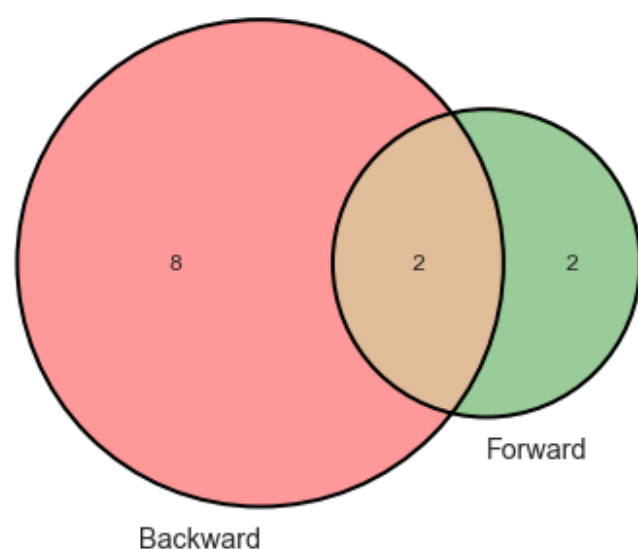
Scopus=s0 /State	Delta Visited	Delta Selected	Delta Precision%	Delta Recall%	Delta F_Measure%	Operation	Accumulated Visited	Accumulated Selected	Accumulated Precision%	Accumulated Recall%	Accumulated F_Measure%
s0: scopus	15	7	46.67	13.73	21.21	-	15	7	46.67	13.73	21.21
s1	46	2	4.35	3.92	4.12	forward	61	9	14.75	17.65	16.07
s2	2	0	0.00	0.00	0.00	forward	63	9	14.29	17.65	15.79
s3	145	8	5.52	15.69	8.16	backward	160	15	9.38	29.41	14.22
s4	75	2	2.67	3.92	3.17	backward	235	17	7.23	33.33	11.89
s5	9	0	0.00	0.00	0.00	backward	244	17	6.97	33.33	11.53
s6	0	0	nan	0.00	0.00	union	292	19	6.51	37.25	11.08

#### Report Strategy 5 - Measures of Strategy (accumulated)

Out[38]:

Strategy	Visited	Selected	Precision %	Recall %	F-Measure %
Scopus + BS // FS	292	19	6.51	37.25	11.08

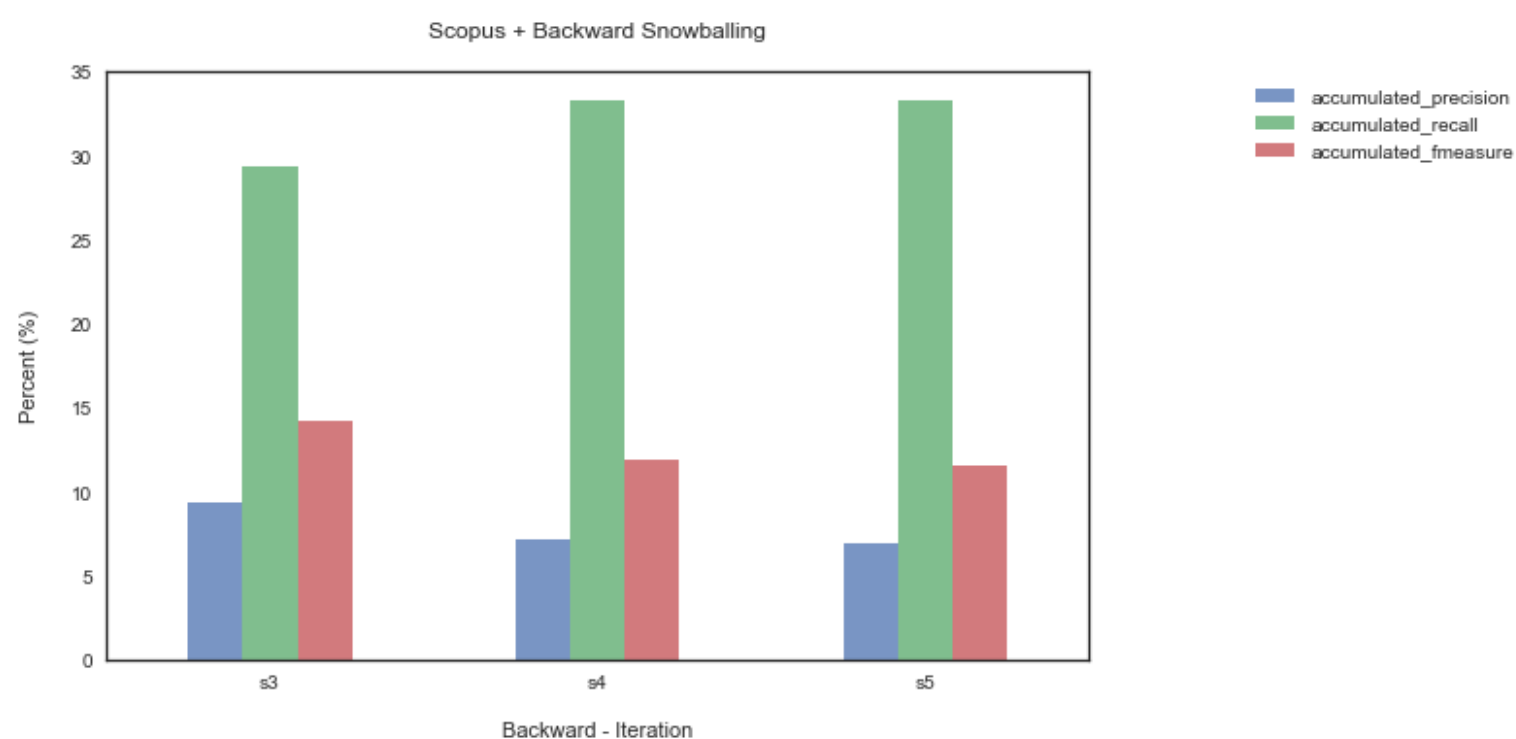
Diagrama de Venn - Strategy 5



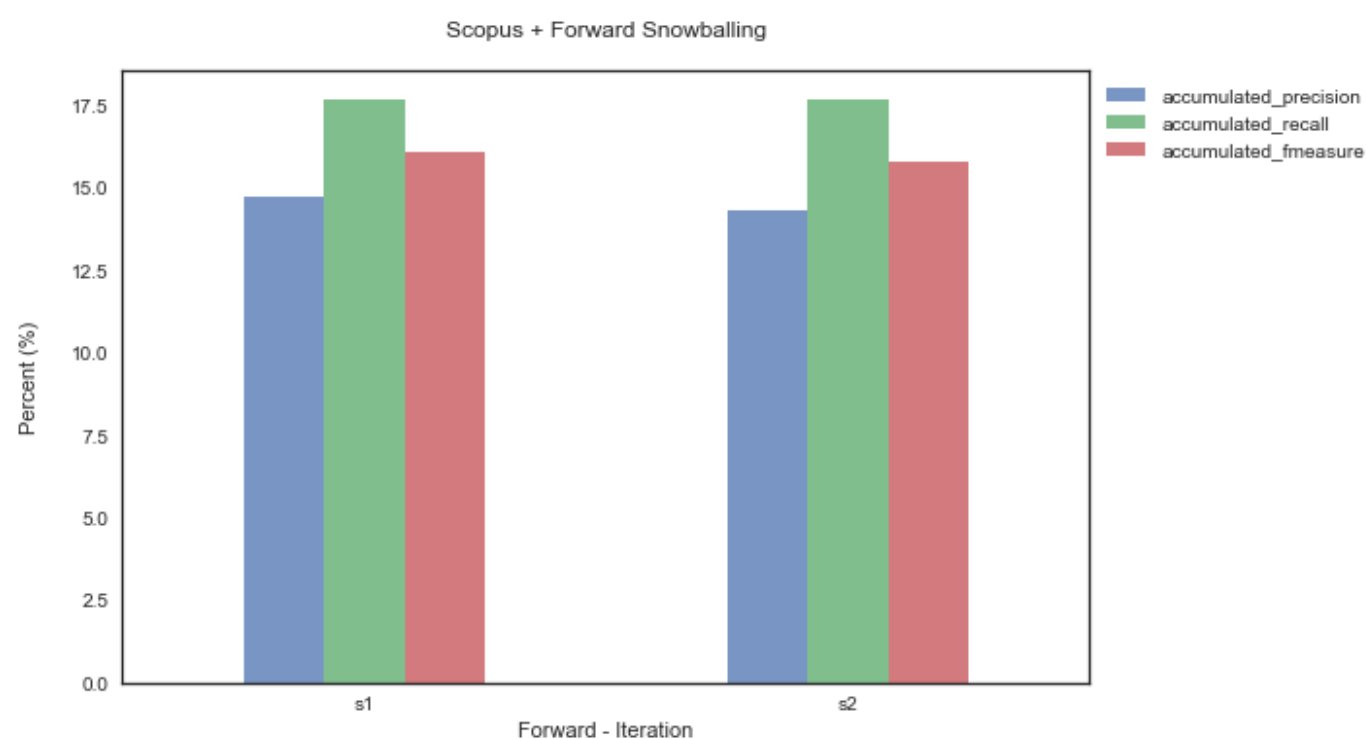
#### Report Diagrama de Venn

Backward ['hinley1995a', 'basili2007a', 'martins2007b', 'trienekens2009a', 'basili2010a', 'mccoy1998a', 'kowalczyk2010a', 'basili2007b', 'liu2005a', 'basili2007c']  
Forward ['basili2010a', 'kowalczyk2010a', 'stallinger2013a', 'petersen2015a']

#### Bar Chart Strategy 5

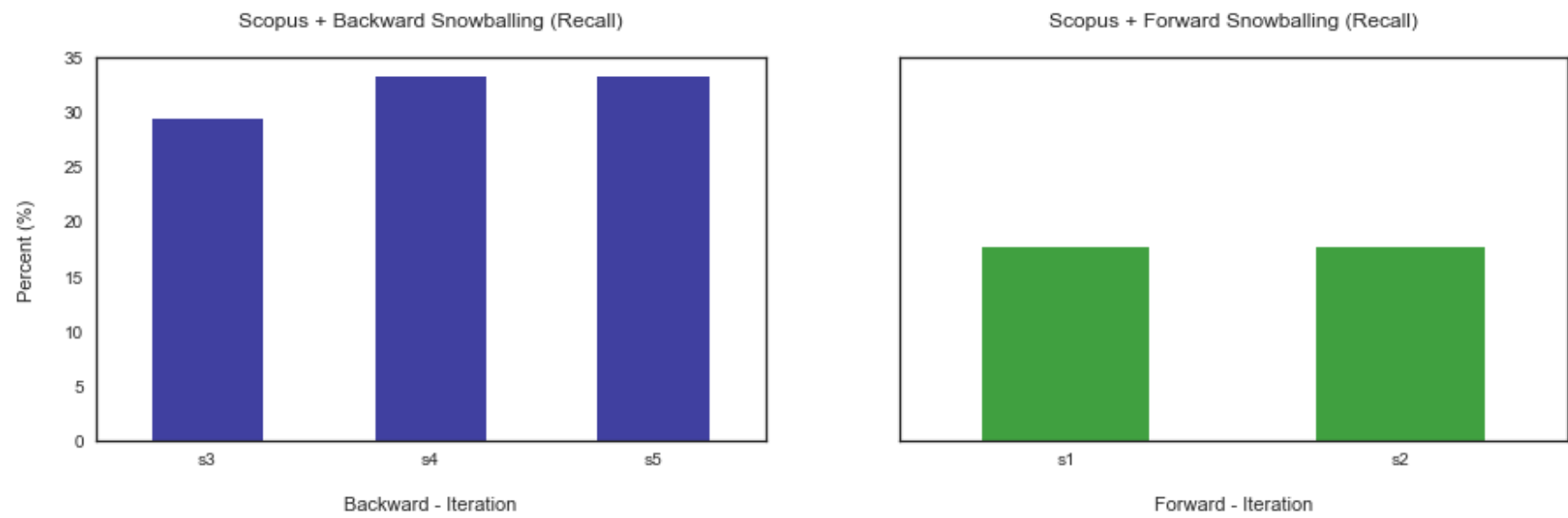


#### Bar Chart Strategy 5



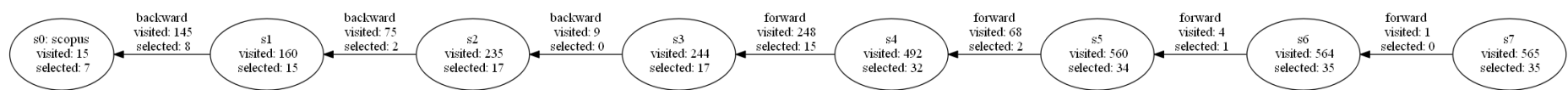


Bar Chart Strategy 5 - (Scopus + BS // FS)



▼ 2.6 Strategy 6 - Scopus + Sequential Snowballing (Scopus + BS + FS)

Provenance Graph - Strategy 6 - Scopus + Sequential Snowballing (Scopus + BS + FS)



Report Strategy 6 - Scopus + Sequential Snowballing (Scopus + BS + FS)

Total of selected studies in the SLR = 51

Precision = Selected/ Visited  
Recall = Selected/ Total of selected studies in the SLR  
F-Measure = Harmonic Mean of Precision and Recall

Out[46]:

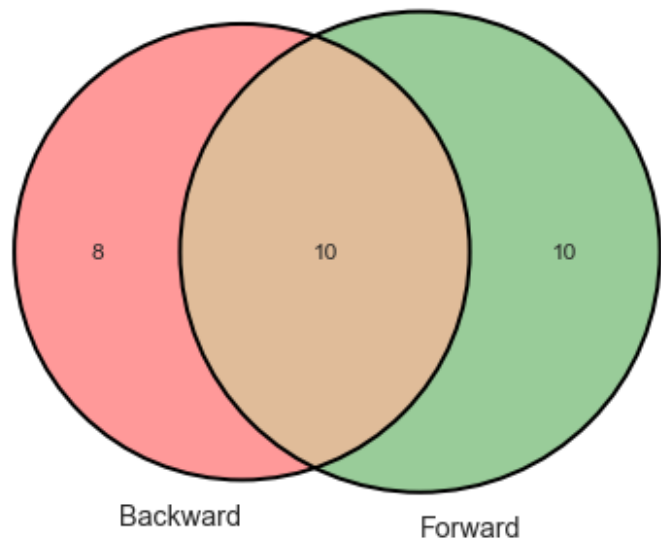
Scopus=s0 /State	Delta Visited	Delta Selected	Delta Precision%	Delta Recall%	Delta F_Measure%	Operation	Accumulated Visited	Accumulated Selected	Accumulated Precision%	Accumulated Recall%	Accumulated F_Measure%
s0: scopus	15	7	46.67	13.73	21.21	-	15	7	46.67	13.73	21.21
s1	145	8	5.52	15.69	8.16	backward	160	15	9.38	29.41	14.22
s2	75	2	2.67	3.92	3.17	backward	235	17	7.23	33.33	11.89
s3	9	0	0.00	0.00	0.00	backward	244	17	6.97	33.33	11.53
s4	248	15	6.05	29.41	10.03	forward	492	32	6.50	62.75	11.79
s5	68	2	2.94	3.92	3.36	forward	560	34	6.07	66.67	11.13
s6	4	1	25.00	1.96	3.64	forward	564	35	6.21	68.63	11.38
s7	1	0	0.00	0.00	0.00	forward	565	35	6.19	68.63	11.36

Report Strategy 6 - Measures of Strategy (accumulated)

Out[47]:

Strategy	Visited	Selected	Precision %	Recall %	F-Measure %
Scopus + BS + FS	565	35	6.19	68.63	11.36

Diagrama de Venn - Strategy 6

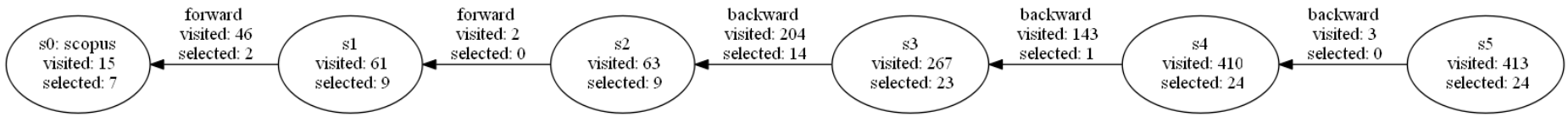


Report Diagrama de Venn

Backward ['hinley1995a', 'basili2007a', 'kowalczyk2010a', 'münch2013a', 'mandić2010a', 'liu2005a', 'trienekens2009a', 'basili2007c', 'martins2007b', 'basili2010a', 'mccoy1998a', 'becker2008b', 'mandić2010b', 'basili2007b', 'trendowicz2011a', 'armbrust2010a', 'mandić2010d', 'münch2013c']  
Forward ['sun2010a', 'kowalczyk2010a', 'münch2013a', 'armbrust2010b', 'trendowicz2014a', 'mandić2010a', 'cocozza2014a', 'stallinger2013a', 'petersen2015a', 'unterkalmsteiner2014a', 'barreto2010a', 'basili2010a', 'birkhölzer2011a', 'becker2008b', 'mandić2010b', 'trendowicz2011a', 'armbrust2010a', 'mandić2010d', 'münch2013c', 'martins2008a']

2.7 Strategy 7 - Scopus + Sequential Snowballing (Scopus + FS + BS)

Provenance Graph - Strategy 7 - Scopus + Sequential Snowballing (Scopus + FS + BS)



Report Strategy 7 - Scopus + Sequential Snowballing (Scopus + FS + BS)

Total of selected studies in the SLR = 51

Precision = Selected/ Visited  
Recall = Selected/ Total of selected studies in the SLR  
F-Measure = Harmonic Mean of Precision and Recall

Out[52]:

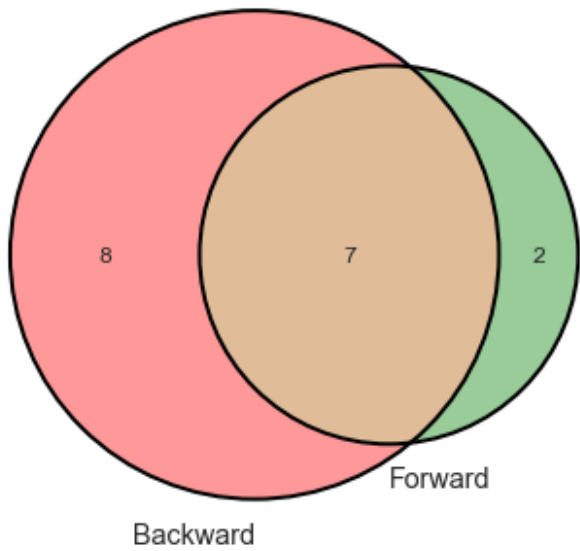
Scopus=s0 /State	Delta Visited	Delta Selected	Delta Precision%	Delta Recall%	Delta F_Measure%	Operation	Accumulated Visited	Accumulated Selected	Accumulated Precision%	Accumulated Recall%	Accumulated F_Measure%
s0: scopus	15	7	46.67	13.73	21.21	-	15	7	46.67	13.73	21.21
s1	46	2	4.35	3.92	4.12	forward	61	9	14.75	17.65	16.07
s2	2	0	0.00	0.00	0.00	forward	63	9	14.29	17.65	15.79
s3	204	14	6.86	27.45	10.98	backward	267	23	8.61	45.10	14.47
s4	143	1	0.70	1.96	1.03	backward	410	24	5.85	47.06	10.41
s5	3	0	0.00	0.00	0.00	backward	413	24	5.81	47.06	10.34

Report Strategy 7 - Measures of Strategy (accumulated)

Out[53]:

Strategy	Visited	Selected	Precision %	Recall %	F-Measure %
Scopus + FS + BS	413	24	5.81	47.06	10.34

Diagrama de Venn - Strategy 7



Report Diagrama de Venn

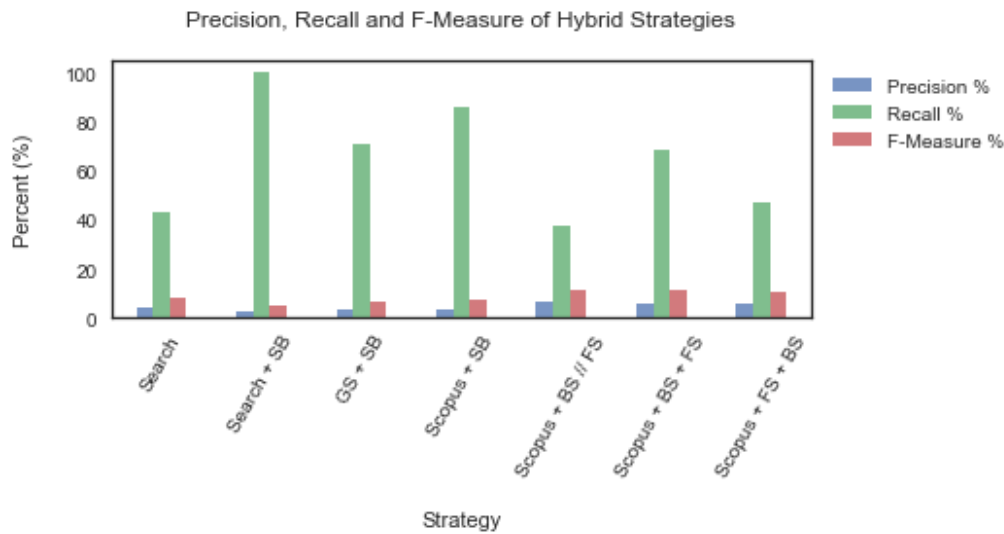
Backward ['hinley1995a', 'basili2007a', 'martins2007b', 'trienekens2009a', 'basili2010a', 'mccoy1998a', 'mandić2010d', 'münch2013c', 'kowalczyk2010a', 'münch2013a', 'mandić2010b', 'basili2007b', 'liu2005a', 'basili2007c', 'trendowicz2011a']  
Forward ['basili2010a', 'mandić2010d', 'münch2013c', 'kowalczyk2010a', 'münch2013a', 'mandić2010b', 'stallinger2013a', 'petersen2015a', 'trendowicz2011a']

3 Analyses

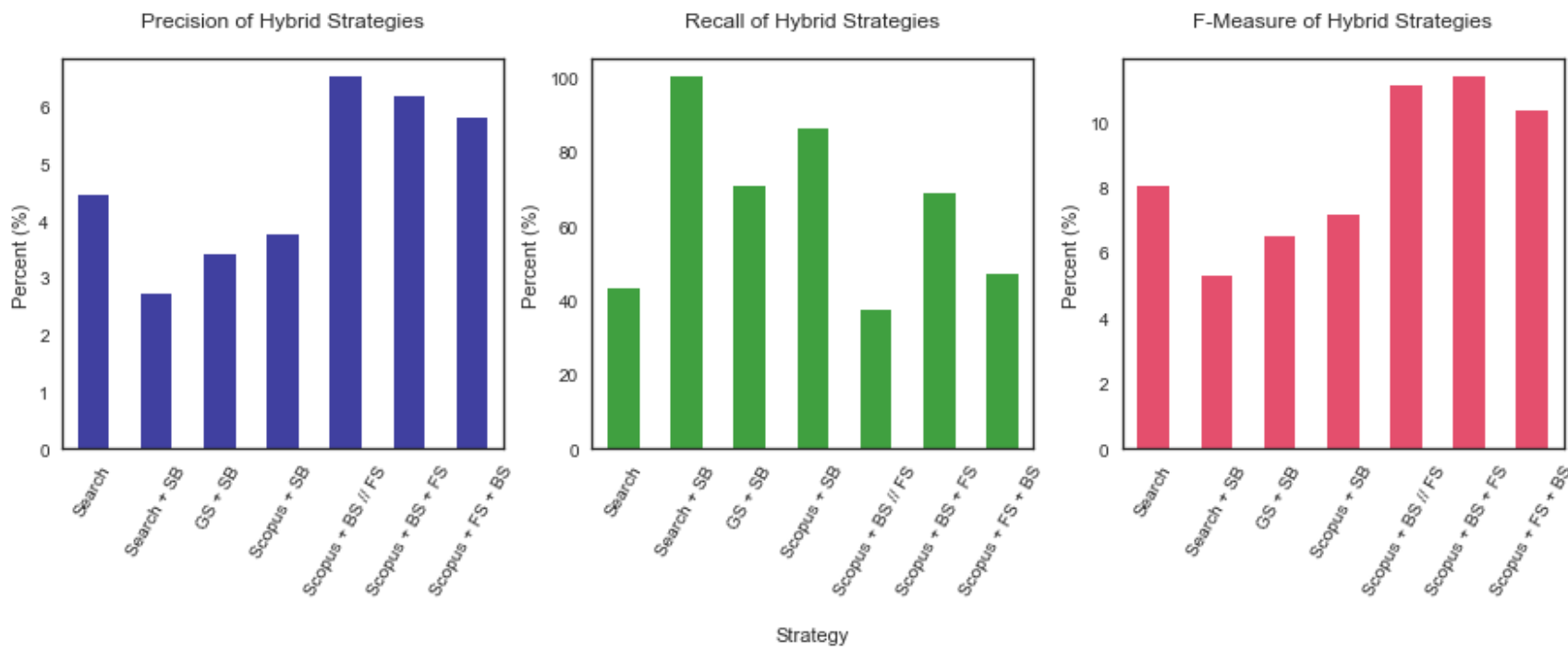
Out[57]:

Strategy	Visited	Selected	Precision %	Recall %	F-Measure %
Search	497	22	4.43	43.14	8.03
Search + SB	1873	51	2.72	100.00	5.30
GS + SB	1059	36	3.40	70.59	6.49
Scopus + SB	1174	44	3.75	86.27	7.18
Scopus + BS // FS	292	19	6.51	37.25	11.08
Scopus + BS + FS	565	35	6.19	68.63	11.36
Scopus + FS + BS	413	24	5.81	47.06	10.34

Bar Chart Strategies



Bar Chart Strategies



▼ 4 Research Questions

RQ1) Which of the digital libraries used in the published SLR is more efficient?

Out[63]:

Digital Library	Visited	Selected	Precision %	Recall %	F-Measure %
acm	100	5	5.00	9.80	6.62
elcompendex	13	5	38.46	9.80	15.62
ieee	43	6	13.95	11.76	12.77
sciencedirect	195	1	0.51	1.96	0.81
scopus	15	7	46.67	13.73	21.21
springer	141	2	1.42	3.92	2.08
webofscience	10	5	50.00	9.80	16.39

RQ2) Which of the hybrid strategies is more efficient?

Out[64]:

Strategy	Visited	Selected	Precision %	Recall %	F-Measure %
Search + SB	1873	51	2.72	100	5.3
GS + SB	1059	36	3.4	70.59	6.49
Scopus + SB	1174	44	3.75	86.27	7.18
Scopus + BS // FS	292	19	6.51	37.25	11.08
Scopus + BS + FS	565	35	6.19	68.63	11.36
Scopus + FS + BS	413	24	5.81	47.06	10.34

RQ3) Which of the strategies is more efficient?

Out[65]:

Strategy	Visited	Selected	Precision %	Recall %	F-Measure %
Search	497	22	4.43	43.14	8.03
Search + SB	1873	51	2.72	100	5.3
GS + SB	1059	36	3.4	70.59	6.49
Scopus + SB	1174	44	3.75	86.27	7.18
Scopus + BS // FS	292	19	6.51	37.25	11.08
Scopus + BS + FS	565	35	6.19	68.63	11.36
Scopus + FS + BS	413	24	5.81	47.06	10.34

RQ4) How many selected articles does the digital library find through the direct search for the title of the article?

Total of selected studies in the SLR = 51

Out[66]:

Digital Library	Selected	Recall %
scopus	42	82.4
springer	16	31.4
ieee	14	27.5
sciencedirect	2	3.9
acm	6	11.8
webofscience	27	52.9
elcompendex	35	68.6