BAB IV
PENGUJIAN

4.1 Pengujian Terhadap Banyak Kasus

Data Abstrak dokumen diujikan dipilih dari setiap klasifikasi dokumen menggunakan teknik sampling sistematik untuk dijadikan sampel yang diujikan menggunakan algoritma *Baum Welch*. Klasifikasi dokumen dapat dilihat pada bab 3 subbab 3.1.1 pada Tabel III.15. Data sampel dokumen abstrak yang dipilih sejumlah 13 abstrak dengan pengujian sebanyak 3 kali pada setiap abstrak. 13 data sampel abstrak yang terdiri dari 5 abstrak dari data abstrak dengan kata kurang dari 150 kata, 4 abstrak dari data abstrak dengan kata dari 150 sampai 200 kata dan 4 abstrak dari data abstrak dengan kata kurang dari 350 kata.

Pengujian dilakukan terhadapt 2 tahap yaitu tahap pertama abstrak yang diujikan tanpa melalui penghilangan kata tidak penting *(stopword)* dan tahap kedua abstrak yang diujikan akan melalui penghilangan kata tidak penting *(stopword)*. Hasil pengujian abstrak tanpa penghilangan kata tidak penting *(stopword)* dapat dilihat pada Tabel IV-1 dan hasil pengujian abstrak dengan tanpa penghilangan kata tidak penting *(stopword)* dapat dilihat pada Tabel IV-2.

Pada kedua tabel hasil pengujian terdapat kolom probabilitas. Probabilitas ini didapat dari hasil pengujian setiap abstrak yang dilakukan sebanyak tiga kali pengujian kemudian menghasilkan beberapa kata kunci. Apabila salah satu kata kunci muncul sebanyak 3 kali maka probabilitas bernilai 100%. Jika kata kunci hanya muncul 1 kali dalam ketiga pengujian itu maka probabilitas bernilai sepertiga dari 100%. Jadi, kolom probabilitas merupakan persentase kemunculan satu kata kunci muncul pada tiga kali pengujian pada abstrak yang diujikan.

Tabel IV.1 Pengujian abstrak tanpa penghilangan kata tidak penting (*stopword*)

Data abstrak ke-	Pengujian ke-	Jumlah kata kunci	Kata kunci yang dihasilkan	Probabilitas
Data sampel abstrak dengan kata kurangdari 150 kata				
2	1	10	*water resources, and resources, and manual, based on, law is, nsgm is, the nsgm, the water, related to, as related*	*water resources :* 100% *and resources :* 100% *and manual :* 100% *based on :* 100% *law is :* 100% *nsgm is :* 100% *the nsgm :* 100% *the water :* 100% *related to :* 100% *as related :* 100%
	2	10	*water resources, and resources, and manual, based on, law is, nsgm is, the nsgm, the water, related to, as related*	
	3	10	*water resources, and resources, and manual, based on, law is, nsgm is, the nsgm, the water, related to, as related*	
4	1	10	*water quality, to the, quality indexes, between the, at the, is the, the water, quality management, quality assessment, and water*	*water quality :* 100% *to the :* 100% *quality indexes :* 100% *between the :* 100% *at the :* 100% *is the :* 100% *the water :* 100% *quality management :* 100% *quality assessment :* 100% *and water :* 100%
	2	10	*water quality, to the, quality indexes, between the, at the, is the, the water, quality management, quality assessment, and water*	
	3	10	*water quality, to the, quality indexes, between the, at the, is the, the water, quality management, quality assessment, and water*	
6	1	10	*enrichment of, of an, study of, growth of, green algae, blue green, that the, of blue, the enrichment, basis of*	*enrichment of :* 100% *of an :* 100% *study of :* 100% *growth of :* 100% *green algae :* 100% *blue green :* 100% *that the :* 100% *of blue :* 100% *the enrichment :* 100% *basis of :* 100%
	2	10	*enrichment of, of an, study of, growth of, green algae, blue green, that the, of blue, the enrichment, basis of*	
	3	10	*enrichment of, of an, study of, growth of, green algae, blue green, that the, of blue, the enrichment, basis of*	
8	1	10	*the seismic, measurement in, the groundwater, in groundwater ,seismic method, reservoir in, depth in, method in, the echo, seismic measurement*	*the seismic :* 100% *measurement in :* 100% *the groundwater :* 100% *in groundwater :* 100% *seismic method :* 100% *reservoir in :* 100% *depth in :* 100% *method in :* 100% *the echo :* 100%
	2	10	*the seismic, measurement in, the groundwater, in groundwater ,seismic method, reservoir in, depth in, method in, the echo,*	

			seismic measurement	*seismic measurement :* 100%
	3	10	*the seismic, measurement in, the groundwater, in groundwater ,seismic method, reservoir in, depth in, method in, the echo, seismic measurement*	
10	1	10	*catchment area, the erosion, the catchment, the tondano, the lake, erosion at, land erosion, area of, the land, lake catchment*	*catchment area :* 100% *the erosion :* 100% *the catchment* : 100% *the tondano :* 100% *the lake :* 100% *erosion at :* 100% *land erosion :* 100% *area of :* 100% *the land :* 100% *lake catchment :* 100%
	2	10	*catchment area, the erosion, the catchment, the tondano, the lake, erosion at, land erosion, area of, the land, lake catchment*	
	3	10	*catchment area, the erosion, the catchment, the tondano, the lake, erosion at, land erosion, area of, the land, lake catchment*	
Data sampel abstrak dengan kata dari 150 sampai 200 kata				
16	1	10	eutrofikasi dan, model yang, di indonesia, yang telah, model eutrofikasi, air yang, yang berfungsi, kajian model, yang dapat, air dan	eutrofikasi dan : 100% model yang : 100% di Indonesia : 100% yang telah : 100% model eutrofikasi : 100% air yang : 100% yang berfungsi : 100% kajian model : 100% yang dapat : 100% air dan : 100%
	2	10	eutrofikasi dan, model yang, di indonesia, yang telah, model eutrofikasi, air yang, yang berfungsi, kajian model, yang dapat, air dan	
	3	10	eutrofikasi dan, model yang, di indonesia, yang telah, model eutrofikasi, air yang, yang berfungsi, kajian model, yang dapat, air dan	
20	1	10	laju erosi, erosi dan, erosi yang, waduk wonogiri, yang paling, dan akhirnya, yang tinggi, aliran permukaan, erosi permukaan, dan aliran	laju erosi : 100% erosi dan : 100% erosi yang : 100% waduk wonogiri : 100% yang paling : 100% dan akhirnya : 100% yang tinggi : 100% aliran permukaan : 100% erosi permukaan : 100% dan aliran : 100%
	2	10	laju erosi, erosi dan, erosi yang, waduk wonogiri, yang paling, dan akhirnya, yang tinggi, aliran permukaan, erosi permukaan, dan aliran	
	3	10	laju erosi, erosi dan, erosi yang, waduk wonogiri, yang paling, dan akhirnya, yang tinggi, aliran permukaan, erosi permukaan, dan aliran	
24	1	10	*recharge of, ground water, of recharge, the nomograph, the change, change of, water recharge, of ground, of natural, estimation of*	*recharge of :* 100% *ground water :* 100% *of recharge :* 100% *the nomograph :* 100% *the change :* 100%

	2	10	*recharge of, ground water, of recharge, the nomograph, the change, change of, water recharge, of ground, of natural, estimation of*	*change of* : 100% *water recharge :* 100% *of ground :* 100% *of natural :* 100% *estimation of :* 100%
	3	10	*recharge of, ground water, of recharge, the nomograph, the change, change of, water recharge, of ground, of natural, estimation of*	
28	1	10	diaktivasi dan, dan adsorben, limbah cair, koagulan dan, adsorben alami, dan limbah, dan zn, dan tanpa, cair elektroplating, cu dan	diaktivasi dan : 100% dan adsorben : 100% limbah cair : 100% koagulan dan : 100% adsorben alami : 100% dan limbah : 100% dan zn : 100% dan tanpa : 100% cair electroplating : 100% cu dan : 100%
	2	10	diaktivasi dan, dan adsorben, limbah cair, koagulan dan, adsorben alami, dan limbah, dan zn, dan tanpa, cair elektroplating, cu dan	
	3	10	diaktivasi dan, dan adsorben, limbah cair, koagulan dan, adsorben alami, dan limbah, dan zn, dan tanpa, cair elektroplating, cu dan	
Data sampel abstrak dengan kata kurang dari 350 kata				
33	1	10	*of the, water supply, in the, of water, supply water, water in, water is, water wheel, irrigation water, pusair water*	*of the* : 100% *water supply :* 100% *in the :* 100% *of water :* 100% *supply water :* 100% *water in :* 100% *water is :* 100% *water wheel* : 100% *irrigation water* : 100% *pusair water* : 100%
	2	10	*of the, water supply, in the, of water, supply water, water in, water is, water wheel, irrigation water, pusair water*	
	3	10	*of the, water supply, in the, of water, supply water, water in, water is, water wheel, irrigation water, pusair water*	
35	1	10	*the damage, the soil, the swelling, expansive soils, the properties, the method, of expansive, of the, the problems, by expansive*	*the damage :* 100% *the soil :* 100% *the swelling* : 100% *expansive soils* : 100% *the properties* : 100% *the method* : 100% *of expansive* : 100% *of the :* 100% *the problems :* 100% *by expansive :* 100%
	2	10	*the damage, the soil, the swelling, expansive soils, the properties, the method, of expansive, of the, the problems, by expansive*	
	3	10	*the damage, the soil, the swelling, expansive soils, the properties, the method, of expansive, of the, the problems, by expansive*	
37	1	10	*citarum river, the citarum, organic waste, of organic, in the, upper citarum, water quality, in citarum, waste in, in upper*	*citarum river :* 100% *the citarum :* 100% *organic waste :* 100% *of organic :* 100%

	2	10	*citarum river, the citarum, organic waste, of organic, in the, upper citarum, water quality, in citarum, waste in, in upper*	*in the :* 100% *upper citarum :* 100% *water quality :* 100% *in citarum :* 100% *waste in :* 100% *in upper :* 100%
	3	10	*citarum river, the citarum, organic waste, of organic, in the, upper citarum, water quality, in citarum, waste in, in upper*	
39	1	10	*maintenance and, the instrumentation, systems and, aspects of, of instrumentation, of the, concepts and, and the, replacement and, and maintenance*	*maintenance and :* 100% *the instrumentation :* 100% *systems and :* 100% *aspects of :* 100% *of instrumentation :* 100% *of the :* 100% *concepts and :* 100% *and the* : 100% *replacement and :* 100% *and maintenance :* 100%
	2	10	*maintenance and, the instrumentation, systems and, aspects of, of instrumentation, of the, concepts and, and the, replacement and, and maintenance*	
	3	10	*maintenance and, the instrumentation, systems and, aspects of, of instrumentation, of the, concepts and, and the, replacement and, and maintenance*	

Tabel IV.2 Pengujian abstrak dengan penghilangan kata tidak penting (*stopword*)

Data abstrak ke-	**Pengujian**	**Jumlah kata kunci**	**Kata kunci yang dihasilkan**	**Probabilitas**
Data sampel abstrak dengan kata kurangdari 150 kata				
2	1	10	*water resources, related water, related nsgm, law related, nsgm based, resources states, water law, nsgm utilizing, law no.7, law having*	*water resources :* 100% *related water :* 100% *related nsgm :* 100% *law related :* 100% *nsgm based :* 100% *resources states :* 100% *water law :* 100% *nsgm utilizing :* 100% *law no.7 :* 100% *law having :* 100%
	2	10	*water resources, related water, related nsgm, law related, nsgm based, resources states, water law, nsgm utilizing, law no.7, law having*	
	3	10	*water resources, related water, related nsgm, law related, nsgm based, resources states, water law, nsgm utilizing, law no.7, law having*	
4	1	10	*water quality, quality indexes, quality conditions, quality assessment, spatial water, quality index, quality management, map water, integration thematic,*	*water quality :* 100% *quality indexes* : 100% *quality conditions :* 100% *quality assessment :*

			thematic map	100%
	2	10	*water quality, quality indexes, quality conditions, quality assessment, spatial water, quality index, quality management, map water, integration thematic, thematic map*	*spatial water* : 100% *quality index :* 100% *quality management :* 100% *map water :* 100% *integration thematic :* 100% *thematic map :* 100%
	3	10	*water quality, quality indexes, quality conditions, quality assessment, spatial water, quality index, quality management, map water, integration thematic, thematic map*	
6	1	0	-	-
	2	0	-	
	3	0	-	
8	1	10	*seismic measurement, seismic method, seismic refraction, measurement groundwater, depth groundwater, echo sounding, artificial vibration, measurement echo, refraction measurement, method groundwater*	*seismic measurement :* 100% *seismic method :* 100% *seismic refraction :* 100% *measurement groundwater* : 100% *depth groundwater :* 100% *echo sounding :* 100% *artificial vibration :* 100% *measurement echo :* 100% *refraction measurement :* 100% *method groundwater :* 100%
	2	10	*seismic measurement, seismic method, seismic refraction, measurement groundwater, depth groundwater, echo sounding, artificial vibration, measurement echo, refraction measurement, method groundwater*	
	3	10	*seismic measurement, seismic method, seismic refraction, measurement groundwater, depth groundwater, echo sounding, artificial vibration, measurement echo, refraction measurement, method groundwater*	
10	1	10	*catchment area, land erosion, lake catchment, tondano lake, area lake, lake tondano, total area, tons/ha/year area, erosion tondano, area studied*	*catchment area :* 100% *land erosion :* 100% *lake catchment :* 100% *tondano lake :* 100% *area lake :* 100% *lake tondano* : 100% *total area :* 100% *tons/ha/year area :* 100% *erosion tondano :* 100% *area studied :* 100%
	2	10	*catchment area, land erosion, lake catchment, tondano lake, area lake, lake tondano, total area, tons/ha/year area, erosion tondano, area studied*	
	3	10	*catchment area, land erosion, lake catchment, tondano lake, area lake, lake tondano, total area, tons/ha/year area, erosion tondano, area studied*	
Data sampel abstrak dengan kata dari 150 sampai 200 kata				

16	1	10	model eutrofikasi, kajian model, eutrofikasi perhatian, konsep model, model matematis, pengendalian eutrofikasi, indonesia kajian, permasalahan eutrofikasi, model konsep, kualitas air	model eutrofikasi : 100% kajian model : 100% eutrofikasi perhatian : 100% konsep model : 100% model matematis : 100% pengendalian eutrofikasi : 100% indonesia kajian : 100% permasalahan eutrofikasi : 100% model konsep : 100% kualitas air : 100%
	2	10	model eutrofikasi, kajian model, eutrofikasi perhatian, konsep model, model matematis, pengendalian eutrofikasi, indonesia kajian, permasalahan eutrofikasi, model konsep, kualitas air	
	3	10	model eutrofikasi, kajian model, eutrofikasi perhatian, konsep model, model matematis, pengendalian eutrofikasi, indonesia kajian, permasalahan eutrofikasi, model konsep, kualitas air	
20	1	10	laju erosi, waduk wonogiri, aliran permukaan, permukaan dianalisis, erosi permukaan, erosi mengendap, erosi aliran, waduk multi, waduk berkurang, pengendali banjir	laju erosi : 100% waduk wonogiri : 100% aliran permukaan : 100% permukaan dianalisis : 100% erosi permukaan : 100% erosi mengendap : 100% erosi aliran : 100% waduk multi : 100% waduk berkurang : 100% pengendali banjir : 100%
	2	10	laju erosi, waduk wonogiri, aliran permukaan, permukaan dianalisis, erosi permukaan, erosi mengendap, erosi aliran, waduk multi, waduk berkurang, pengendali banjir	
	3	10	laju erosi, waduk wonogiri, aliran permukaan, permukaan dianalisis, erosi permukaan, erosi mengendap, erosi aliran, waduk multi, waduk berkurang, pengendali banjir	
24	1	10	*ground water, water recharge, artificial recharge, aquifer recharge, estimation recharge, rain water, water sources, water exploitation, increase ground, recharge nomograph*	*ground water* : 100% *water recharge :* 100% *artificial recharge :* 100% *aquifer recharge* : 100% *estimation recharge :* 100% *rain water :* 100% *water sources :* 100% *water exploitation :* 100% *increase ground :* 100% *recharge nomograph :* 100%
	2	10	*ground water, water recharge, artificial recharge, aquifer recharge, estimation recharge, rain water, water sources, water exploitation, increase ground, recharge nomograph*	
	3	10	*ground water, water recharge, artificial recharge, aquifer recharge, estimation recharge, rain water, water sources, water*	

			exploitation, increase ground, recharge nomograph	
28	1	10	limbah cair, diaktivasi limbah, koagulan adsorben, adsorben alami, diaktivasi azolla, cair elektroplating, diaktivasi adsorben, limbah teh, limbah elektroplating, azolla diaktivasi	limbah cair : 100% diaktivasi limbah : 100% koagulan adsorben : 100% adsorben alami : 100% diaktivasi azolla : 100% cair electroplating : 100% diaktivasi adsorben : 100% limbah the : 100% limbah electroplating : 100% azolla diaktivasi : 100%
	2	10	limbah cair, diaktivasi limbah, koagulan adsorben, adsorben alami, diaktivasi azolla, cair elektroplating, diaktivasi adsorben, limbah teh, limbah elektroplating, azolla diaktivasi	
	3	10	limbah cair, diaktivasi limbah, koagulan adsorben, adsorben alami, diaktivasi azolla, cair elektroplating, diaktivasi adsorben, limbah teh, limbah elektroplating, azolla diaktivasi	
Data sampel abstrak dengan kata kurang dari 350 kata				
33	1	10	*supply water, water supply, irrigation water, pusair water, paddy production, water wheel, wheel supply, dependable water, long water, problem water*	*supply water :* 100% *water supply :* 100% *irrigation water :* 100% *pusair water :* 100% *paddy production :* 100% *water wheel :* 100% *wheel supply :* 100% *dependable water :* 100% *long water :* 100% *problem water :* 100%
	2	10	*supply water, water supply, irrigation water, pusair water, paddy production, water wheel, wheel supply, dependable water, long water, problem water*	
	3	10	*supply water, water supply, irrigation water, pusair water, paddy production, water wheel, wheel supply, dependable water, long water, problem water*	
35	1	9	*expansive soils, expansive soil, expansive, damage caused, caused expansive, swelling soils, soils widely, swelling soil, soil movement*	*expansive soils :* 100% *expansive soil :* 100% *expansive :* 100% *damage caused :* 100% *caused expansive :* 100% *swelling soils :* 100% *soils widely :* 100% *swelling soil :* 100% *soil movement :* 100%
	2	9	*expansive soils, expansive soil, expansive, damage caused, caused expansive, swelling soils, soils widely, swelling soil, soil movement*	
	3	9	*expansive soils, expansive soil, expansive, damage caused, caused expansive, swelling soils, soils widely, swelling soil, soil movement*	
37	1	10	*citarum river, organic waste, upper citarum, water quality, river polluted, research organic, farm waste, sources organic,*	*citarum river :* 100% *organic waste :* 100% *upper citarum :* 100% *water quality :* 100%

			domestic waste, waste upper	*river polluted :* 100% *research organic :* 100% *farm waste :* 100% *sources organic :* 100% *domestic waste :* 100% *waste upper :* 100%
	2	10	*citarum river, organic waste, upper citarum, water quality, river polluted, research organic, farm waste, sources organic, domestic waste, waste upper*	
	3	10	*citarum river, organic waste, upper citarum, water quality, river polluted, research organic, farm waste, sources organic, domestic waste, waste upper*	
39	1	10	*instrument systems, aspects instrumentation, instrumentation program, replacement maintenance, instrumentation including, systems continuous, instrumentation planning, maintenance rehabilitation, design maintenance, modification replacement*	*instrument systems :* 100% *aspects instrumentation :* 100% *instrumentation program* : 100% *replacement maintenance :* 100% *instrumentation including* : 100% *systems continuous :* 100% *instrumentation planning* : 100% *maintenance rehabilitation :* 100% *design maintenance :* 100% *modification replacement* : 100%
	2	10	*instrument systems, aspects instrumentation, instrumentation program, replacement maintenance, instrumentation including, systems continuous, instrumentation planning, maintenance rehabilitation, design maintenance, modification replacement*	
	3	10	*instrument systems, aspects instrumentation, instrumentation program, replacement maintenance, instrumentation including, systems continuous, instrumentation planning, maintenance rehabilitation, design maintenance, modification replacement*	

Berdasarkan hasil pengujian dari 13 sampel abstrak yang diujikan, sebagian besar kata kunci menghasilkan 10 jumlah kata kunci. Adapun satu abstrak yang menghasillkan 9 kata kunci, bahkan tidak menghasilkan kata kunci sama sekali. Kata kunci yang dibutuhkan adalah 5 kata kunci, maka kata kunci yang akan diambil dan digunakan sebagai pembanding dari setiap abstrak adalah 5 kata kunci dari urutan kemunculan pertama atau teratas.

5 kata kunci yang dihasilkan akan dibandingkan dengan kata kunci yang telah ada dan menghasilkan nilai persentase kesamaan kata kunci baru atau yang dihasilkan dari percobaan dengan kata kunci awal atau yang telah ada.

Perhitungan nilai persentasi kesamaan kata kunci untuk setiap dokumen abstrak dapat dihitung dengan cara sebagai berikut:

Persentase setiap kata kunci yang dihasilkan diambil dari membandingkan kesamaan dari hasil kata kunci dengan kata kunci yang telah ada. Satu persentase kata kunci akan dihitung dengan cara sebagai berikut:

$$\text{Nilai persentase 1 kata kunci } = \frac{1\ jumlah\ kata\ kunci\ awal}{jumlah\ seluruh\ kata\ kunci\ awal}\ x\ 100\%$$

Apabila terdapat kata kunci baru yang sama dengan kata kunci lama maka nilai persentase senilai dengan nilai persentase 1 kata kunci. Apabila terdapat beberapa kata kunci baru yang salah satu katanya sama dengan kata kunci yang lama maka persentase senilai dengan setengah dari persentase harapan satu kata kunci.

Persentase kesamaan kata kunci lama dan kata kunci baru atau kata kunci yang dihasilkan dapat dilihat pada tabel IV.3 dan tabel IV.4.

Tabel IV.3 Persentase kesamaan kata kunci lama dan baru terhadap data sampel abstrak tanpa mengalami penghilangan kata tidak penting *(stopword)*

Data abstrak ke-	Kata Kunci Awal	Kata Kunci Baru	Kesamaan kata kunci	persetase
Data sampel abstrak dengan kata kurangdari 150 kata				
2	*1.standardization* *2.regulation* *3.construction management* *4.water resources management* ∑kata = 4	*1.water resources* *2.and resources* *3.and manual* *4.based on* *5.law is* ∑kata = 5	1. *water resources, and resources* ∑kata = 0,5	12,5% $\left(\frac{1}{4}x100\%\right)$ 0,5 =12,5%
4	*1.thematic map* *2.spatial environment* *3.physical-chemical index* *4.bmwp score* *5.the upper citarum* ∑kata = 5	*1.water quality* *2.to the* *3.quality indexes* *4.between the* *5.at the* ∑kata = 5	- ∑kata = 0	0% $\left(\frac{1}{5}x100\%\right)$ 0 = 0%
6	*1.blue green algae* *2.blue green algae bloom* *3.mineral nutrient* ∑kata = 3	*1.enrichment of* *2.of an* *3.study of* *4.growth of* *5.green algae* ∑kata = 5	*1.green algae* ∑kata = 0,5	16,5% $\left(\frac{1}{3}x100\%\right)$ 0,5 =16,5%
8	*1.seismic* *2.groundwater* *3.river*	*1.the seismic* *2.measurement in* *3.the groundwater*	*1.the seismic* *2.the groundwater*	50%

	4.reservoir ∑kata = 4	*4.in groundwater* *5.seismic method* ∑kata = 5	 ∑kata =2	$\left(\frac{1}{4}x100\%\right)$ 2 = 50%
10	*1.land erosion* *2.usle formula* *3.geographic information system* *4.lake catchment area* ∑kata = 4	*1.catchment area,* *2.the erosion,* *3.the catchment* *4.the tondano* *5.the lake* ∑kata = 5	*1.catchment area,the catchment,the lake* *2.the erosion* ∑kata = 1	25% $\left(\frac{1}{4}x100\%\right)$ 1 = 25%
Rata-rata persentase				20,8%
Data sampel abstrak dengan kata dari 150 sampai 200 kata				
16	1.eutrofikasi 2.model dinamik 3.parameter kualitas air 4.reactor tercampur sempurna 5.model konseptual ∑kata = 5	1.eutrofikasi dan, 2.model yang, 3.di indonesia, 4.yang telah, 5.model eutrofikasi ∑kata = 5	1.eutrofikasi dan,model eutrofikasi 2.model yang,model eutrofikasi ∑kata = 1	20% $\left(\frac{1}{5}x100\%\right)$ 1 = 20%
20	1.reboisasi 2.aliran permukaan 3.erosi 4.sedimentasi 5.metode usle 6.lahan kritis 7.umur guna ∑kata = 7	1.laju erosi 2.erosi dan 3.erosi yang 4.waduk wonogiri 5.yang paling ∑kata = 5	1.laju erosi,erosi dan,erosi yang ∑kata = 1	14,3% $\left(\frac{1}{7}x100\%\right)$ 1 = 14,3%
24	*1.deep well* *2.confined aquifer* *3.artificial recharge* ∑kata = 3	*1.recharge of* *2.ground water* *3.of recharge* *4.the nomograph* *5.the change* ∑kata =5	*1.recharge of,of recharge* ∑kata = 0,5	16,5% $\left(\frac{1}{3}x100\%\right)$ 0,5 =16,5%
28	1.adsorben alami 2.efektivitas koagulan 3.logam berat ∑kata = 3	1.diaktivasi dan 2.dan adsorben 3.limbah cair 4.koagulan dan 5.adsorben alami ∑kata = 5	1. adsorben alami 2.koagulan dan ∑kata = 1,5	50% $\left(\frac{1}{3}x100\%\right)$ 1,5 =50%
Rata-rata persentase				25,2%
Data sampel abstrak dengan kata kurang dari 350 kata				
33	*1.pusair water wheel* *2.water supply* *3.irrigation* *4.paddy production increase* ∑kata = 4	*1.of the* *2.water supply* *3.in the* *4.of water* *5.supply water* ∑kata = 5	*1.water supply* ∑kata = 1	25% $\left(\frac{1}{4}x100\%\right)$ 1 = 25%
35	*1.expansive soil*	*1.the damage*	*1.expansive soil*	25%

	2.swelling *3.moisture change* *4.structures* ∑kata = 4	*2.the soil* *3.the swelling* *4.expansive soil* *5.the properties* ∑kata = 5	 ∑kata = 1	$\left(\frac{1}{4}x100\%\right)1 =$ 25%
37	*1.organic waste* *2.water quality* *3.chemical-physical index* *4.upper citarum river* *5.macrobenthos* ∑kata = 5	*1.citarum river* *2.the citarum* *3.organic waste* *4.of organic* *5.in the* ∑kata = 5	*1.organic waste* *2.citarum river* ∑kata = 1,5	30% $\left(\frac{1}{5}x100\%\right)1{,}5 = 30\%$
39	*1.instrumentation* *2.geotechnical* *3.monitoring* *4.replacement and maintenance* *5.foundations* *6.embankment dams* ∑kata = 6	*1.maintenance and,* *2.the instrumentation,* *3.systems and,* *4.aspects of,* *5.of instrumentation* ∑kata = 5	*1.the instrumention* *2.maintance and* ∑kata = 1,5	25% $\left(\frac{1}{6}x100\%\right)1{,}5$ =25%
Rata-rata persentase				26,3%

Tabel IV.4 Persentase kesamaan kata kunci lama dan baru terhadap data sampel abstrak dengan penghilangan kata tidak penting *(stopword)*

Data abstrak ke-	Kata kunci awal	Kata Kunci Baru	Kesamaan kata kunci	Persentase
Data sampel abstrak dengan kata kurangdari 150 kata				
2	*1.standardization* *2.regulation* *3.construction management* *4.water resources management* ∑kata = 4	*1.water resources* *2.related water* *3.related nsgm* *4.law related* *5.nsgm based* ∑kata = 5	*1.water resource, related water* ∑kata = 0,5	12,5% $\left(\frac{1}{4}x100\%\right)0{,}5$ =12,5%
4	*1.thematic map* *2.spatial environment* *3.physical-chemical index* *4.bmwp score* *5.the upper citarum* ∑kata = 5	*1.water quality* *2.quality indexes* *3.quality conditions* *4.quality assessment* *5.spatial water* ∑kata = 5	- ∑kata = 0	0% $\left(\frac{1}{5}x100\%\right)0 =$ 0%
6	*1.blue green algae* *2.blue green algae bloom* *3.mineral nutrient*	-	-	0%

	∑kata = 3	∑kata = 0	∑kata = 0	$\left(\frac{1}{3}x100\%\right)0 =$ 0%
8	*1.seismic* *2.groundwater* *3.river* *4.reservoir* ∑kata = 4	*1.seismic measurement* *2.seismic method* *3.seismic refraction* *4.measurement groundwater* *5.depth groundwater* ∑kata = 5	*1.seismic measurement,seismic method,seismic refraction* *2.measurement groundtwater,depth groundwater* ∑kata = 1	25% $\left(\frac{1}{4}x100\%\right)1 =$ 25%
10	*1.land erosion* *2.usle formula* *3.geographic information system* *4.lake catchment area* ∑kata = 4	*1.catchment area* *2.land erosion* *3.lake catchment* *4.tondano lake* *5.area lake* ∑kata = 5	*1.land erosion* *2.catchment area,lake catchment,area lake* ∑kata = 1,5	37,5% $\left(\frac{1}{4}x100\%\right)1{,}5 = 37{,}5\%$
Rata-rata persentase				15%
Data sampel abstrak dengan kata dari 150 sampai 200 kata				
16	1.eutrofikasi 2.model dinamik 3.parameter kualitas air 4.reactor tercampur sempurna 5.model konseptual ∑kata = 5	1.model eutrofikasi 2.kajian model 3.eutrofikasi perhatian 4.konsep model 5.model matematis ∑kata = 5	1.model eutrofikasi, eutrofikasi perhatian 2.konsep model,model matematis,kajian model ∑kata = 1	20% $\left(\frac{1}{4}x100\%\right)1 =$ 20%
20	1.reboisasi 2.aliran permukaan 3.erosi 4.sedimentasi 5.metode usle 6.lahan kritis 7.umur guna ∑kata = 7	1.laju erosi 2.waduk wonogiri 3.aliran permukaan 4.permukaan dianalisis 5.erosi permukaan ∑kata = 5	1.aliran permukaan 2.laju erosi,erosi permukaan ∑kata = 1,5	21,4% $\left(\frac{1}{7}x100\%\right)1{,}5 =$21,4%
24	*1.deep well* *2.confined aquifer,* *3.artificial recharge* ∑kata = 3	*1.ground water* *2.water recharge* *3.artificial recharge* *4.aquifer recharge* *5.estimation recharge* ∑kata = 5	*1.aquifer recharge* *2.artificial recharge* ∑kata = 1,5	50% $\left(\frac{1}{3}x100\%\right)1{,}5 =$50%

28	1.adsorben alami 2.efektivitas koagulan 3.logam berat ∑kata = 3	1.limbah cair 2.diaktivasi limbah 3.koagulan adsorben 4.adsorben alami 5.diaktivasi azolla ∑kata = 5	1.adsorben alami 2.koagulan adsorben ∑kata = 1,5	50% $\left(\frac{1}{4}x100\%\right)$ 1,5 =12,5%
Rata-rata persentase				35,4%
Data sampel abstrak dengan kata kurang dari 350 kata				
33	*1.pusair water wheel* *2.water supply* *3.irrigation* *4.paddy production increase* ∑kata = 4	*1.supply water* *2.water supply* *3.irrigation water* *4.pusair water* *5.paddy production* ∑kata = 5	*1.pusair water* *2.water supply* *3.irrigation water* *4.paddy production* ∑kata = 2,5	62,5% $\left(\frac{1}{4}x100\%\right)$ 2,5 =62,5%
35	*1.expansive soil* *2.swelling* *3.moisture change* *4.structures* ∑kata = 4	*1.expansive soils* *2.expansive soil* *3.expansive* *4.damage caused* *5.caused expansive* ∑kata = 5	*1.expansive soil* ∑kata = 1	25% $\left(\frac{1}{4}x100\%\right)$ 1 = 25%
37	*1.organic waste* *2.water quality* *3.chemical-physical index* *4.upper citarum river* *5.macrobenthos* ∑kata = 5	*1.citarum river* *2.organic waste* *3.upper citarum* *4.water quality* *5.river polluted* ∑kata = 5	*1.organic waste* *2.water quality* *3.citarum river, upper citarum* ∑kata = 2,5	50% $\left(\frac{1}{5}x100\%\right)$ 2,5 =50%
39	*1.instrumentation* *2.geotechnical* *3.monitoring* *4.replacement and maintenance* *5.foundations* *6.embankment dams* ∑kata = 6	*1.instrument systems* *2.aspects instrumentation,* *3.instrumentation program* *4.replacement maintenance,* *5.instrumentation including* ∑kata = 5	*1.aspects instrumentation, instrumentation program,instrumentation including* *2.replacement maintenance* ∑kata =1	16,7% $\left(\frac{1}{6}x100\%\right)$ 1 = 16,7%
Rata-rata persentase				38,5%

Berdasarkan hasil dari tabel IV.3 dan IV.4, rata-rata persentasi dari setiap klasifikasi dokumen dapat dilihat pada Tabel IV.5 :

Tabel IV.5 Persentase kesamaan kata kunci yang dihasilkan dengan kata kunci awal

Jumlah kata abstrak	**Persentase kesamaan kata kunci awal dengan kata kunci yang dihasilkan**
Tanpa penghilangan kata tidak penting	
<150 kata	20,8%
150-200 kata	25,2%
<350 kata	26,3%
Dengan penghilangan kata tidak penting	
<150 kata	15%
150-200 kata	35,4%
<350 kata	38,5%

Kesesuaian kata kunci dengan kriteria kata kunci yaitu gabungan dari kata benda atau kata sifat serta kaitan dengan judul dan isi abstrak dapat dilihat dari nilai persentasi dari setiap data abstrak pada tabel IV.6 dan tabel IV.7:

Tabel IV.6 Persentasi hasil kesuaian kriteria kata kunci pada setiap data abstrak tanpa penghilangan kata tidak penting

Data abstrak ke-	**Kata Kunci Baru**	**Rangkaian kesesuaian kata kunci**	**Persentase**
2	*1.water resources* *2.and resources* *3.and manual* *4.based on* *5.law is* ∑kata = 5	*1.water resources* ∑kata = 1	20% $\left(\frac{1}{5}x100\%\right)1 = 20\%$
4	*1.water quality* *2.to the* *3.quality indexes* *4.between the* *5.at the* ∑kata = 5	*1.water quality* *2.quality indexes* ∑kata =2	40% $\left(\frac{1}{5}x100\%\right)2 = 40\%$
6	*1.enrichment of* *2.of an* *3.study of* *4.growth of* *5.green algae* ∑kata = 0	- ∑kata = 0	0%
8	*1.the seismic* *2.measurement in* *3.the groundwater* *4.in groundwater*	*1.the seismic* *2.the groundwater* *3. seismic method*	60%

	5.seismic method ∑kata = 5	 ∑kata = 3	 $\left(\frac{1}{5}x100\%\right)3 = 60\%$
10	*1.catchment area,* *2.the erosion,* *3.the catchment,* *4.the tondano,* *5.the lake* ∑kata = 5	*1.catchment area* *2. .the erosion,* ∑kata = 2	40% $\left(\frac{1}{5}x100\%\right)2 = 40\%$
16	1.eutrofikasi dan, 2.model yang, 3.di indonesia, 4.yang telah, 5.model eutrofikasi ∑kata = 5	1.model eutrofikasi ∑kata = 1	80% $\left(\frac{1}{5}x100\%\right)1 = 20\%$
20	1.laju erosi 2.erosi dan 3.erosi yang 4.waduk wonogiri 5.yang paling ∑kata = 5	1.laju erosi 2.waduk wonogiri ∑kata = 2	40% $\left(\frac{1}{5}x100\%\right)2 = 40\%$
24	*1.recharge of* *2.ground water* *3.of recharge* *4.the nomograph* *5.the change* ∑kata = 5	*1.ground water* *2.the nomograph* ∑kata = 2	40% $\left(\frac{1}{5}x100\%\right)2 = 40\%$
28	1.diaktivasi dan 2.dan adsorben 3.limbah cair 4.koagulan dan 5.adsorben alami ∑kata = 5	1.limbah cair 2.adsorben alami ∑kata = 2	40% $\left(\frac{1}{5}x100\%\right)2 = 40\%$
33	*1.of the* *2.water supply* *3.in the* *4.of water* *5.supply water* ∑kata = 5	*1.water supply* ∑kata = 1	20% $\left(\frac{1}{5}x100\%\right)1 = 20\%$
35	*1.the damage* *2.the soil* *3.the swelling* *4.expansive soils* *5.the properties* ∑kata = 5	*1.expansive soil* *2. the swelling* ∑kata = 2	40% $\left(\frac{1}{5}x100\%\right)2 = 40\%$

37	*1.citarum river* *2.the citarum* *3.organic waste* *4.of organic* *5.in the* ∑kata = 5	*1.citarum river* *2.organic waste* ∑kata 2	40% $\left(\frac{1}{5}x100\%\right)2 = 40\%$
39	*1.maintenance and,* *2.the instrumentation,* *3.systems and,* *4.aspects of,* *5.of instrumentation* ∑kata = 5	*1. the instrumentation* ∑kata =1	20% $\left(\frac{1}{5}x100\%\right)1 = 20\%$
Rata-rata persentase			32,3%

Tabel IV.7 Persentasi hasil kesuaian kriteria kata kunci pada setiap data abstrak dengan penghilangan kata tidak penting

Data abstrak ke-	**Kata Kunci Baru**	**Rangkaian kesesuaian kata kunci**	**Persentase**
2	*1.water resources* *2.related water* *3.related nsgm* *4.law related* *5.nsgm based* ∑kata = 5	*1.water resources* *2.related water* *3.related nsgm* ∑kata = 3	60% $\left(\frac{1}{5}x100\%\right)3$ =60%
4	*1.water quality* *2.quality indexes* *3.quality conditions* *4.quality assessment* *5.spatial water* ∑kata = 5	*1.water quality* *2.quality indexes* *3.quality conditions* *4.quality assessment* *5.spatial water* ∑kata = 5	100% $\left(\frac{1}{5}x100\%\right)5$ =100%
6	- ∑kata = 0	- ∑kata = 0	0%
8	*1.seismic measurement* *2.seismic method* *3.seismic refraction* *4.measurement groundwater* *5.depth groundwater* ∑kata = 5	*1.seismic measurement* *2.seismic method* *3.seismic refraction* *5.depth groundwater* ∑kata = 4	80% $\left(\frac{1}{5}x100\%\right)4$ =80%
10	*1.catchment area* *2.land erosion* *3.lake catchment* *4.tondano lake* *5.area lake* ∑kata = 5	*1.catchment area* *2.land erosion* *3.lake catchment* *4.tondano lake* ∑kata = 4	80% $\left(\frac{1}{4}x100\%\right)4 = 80\%$

16	1.model eutrofikasi 2.kajian model 3.eutrofikasi perhatian 4.konsep model 5.model matematis ∑kata = 5	1.model eutrofikasi 2.kajian model 3.konsep model 4.model matematis ∑kata = 4	80% $\left(\frac{1}{5}x100\%\right)4 = 80\%$
20	1.laju erosi 2.waduk wonogiri 3.aliran permukaan 4.permukaan dianalisis 5.erosi permukaan ∑kata = 5	1.laju erosi 2.waduk wonogiri 3.aliran permukaan ∑kata = 3	60% $\left(\frac{1}{5}x100\%\right)3 = 60\%$
24	*1.ground water* *2.water recharge* *3.artificial recharge* *4.aquifer recharge* *5.estimation recharge* ∑kata = 5	*1.ground water* *2.water recharge* *3.artificial recharge* *4.aquifer recharge* *5.estimation recharge* ∑kata = 5	100% $\left(\frac{1}{5}x100\%\right)5 = 100\%$
28	1.limbah cair 2.diaktivasi limbah 3.koagulan adsorben 4.adsorben alami 5.diaktivasi azolla ∑kata = 5	1.limbah cair 3.koagulan adsorben 4.adsorben alami ∑kata = 3	60% $\left(\frac{1}{5}x100\%\right)3 = 60\%$
33	*1.supply water* *2.water supply* *3.irrigation water* *4.pusair water* *5.paddy production* ∑kata = 5	*1.water supply* *2.irrigation water* *4.paddy production* ∑kata = 3	80% $\left(\frac{1}{5}x100\%\right)3 = 80\%$
35	*1.expansive soils* *2.expansive soil* *3.expansive* *4.damage caused* *5.caused expansive* ∑kata = 5	*1.expansive soil* ∑kata = 1	20% $\left(\frac{1}{5}x100\%\right)1 = 20\%$
37	*1.citarum river* *2.organic waste* *3.upper citarum* *4.water quality* *5.river polluted* ∑kata = 5	*1.citarum river* *2.organic waste* *3.upper citarum* *4.water quality* ∑kata 4	80% $\left(\frac{1}{5}x100\%\right)4 = 50\%$
39	*1.instrument systems* *2.aspects instrumentation* *3.instrumentation*	*1.instrument systems* *2.aspects instrumentation*	40%

	program *4.replacement maintenance* *5.instrumentation including* ∑kata = 5	 ∑kata =2	 $\left(\frac{1}{5}x100\%\right)$ 2 =40%
Rata-rata persentase			64,6%

Nilai persentasi pada tabel IV.6 dan tabel IV.7 didapat dari 100% dibagi jumlah hasil kata kunci kemudian dikalikan dengan jumlah kata kunci yang sesuai. Kemudian rata-rata persentasi dari keseluruhan data abstrak adalah jumlah persentasi seluruh data abstrak dibagi dengan jumlah seluruh data abstrak.

Pada setiap pengujian memiliki waktu eksekusi yang berbeda-beda dikarenakan jumlah dari inputan yang diberikan pun berbeda-beda. Dapat dilihat pada table IV.8 waktu yang dibutuhkan pada setiap pengujian data abstrak. Waktu pengujian 1 merupakan waktu yang dibutuhkan tanpa ada penghilangan kata tidak penting *(stopword),* sedangkan waktu pengujian 2 merupakan waktu yang dibutuhkan dengan adanya penghilangan kata tidak penting *(stopword).*

Tabel IV.8 Penghitungan waktu pengujian

Klasifikasi jumlah kata abstrak	**Data abstrak ke-**	**Waktu pengujian 1 (detik)**	**Waktu pengujian 2 (detik)**
<150 kata	2	0.172	0.093
	4	0.141	0.109
	6	0.609	2.371
	8	0.405	0.296
	10	0.437	0.281
150-200 kata	16	0.983	0.406
	20	1.669	0.359
	24	0.936	0.452
	28	1.295	0.499
<350 kata	33	1.092	0.374
	35	1.638	0.89
	37	1.154	0.624
	39	1.763	0.936
Rata-rata waktu		0.945692	0.591538

4.2.1 Kesimpulan Pengujian

Berdasarkan hasil pengujian yang telah dilakukan pada data sampel dokumen abstrak yang diambil menggunakan teknik sampling sistematik, maka pada pengujian abstrak tanpa penghilangan kata tidak penting (*stopword*) memiliki persentase lebih kecil dibandingkan pencarian kata kunci dengan mengalami penghilangan kata tidak penting *(stopword)* terlebih dahulu. Hal ini membuktikan bahwa banyaknya kata tidak penting *(stopword)* dapat mempengaruhi kata kunci yang dihasilkan.

Nilai hasil kata kunci yang mengalami penghilangan kata tidak penting *(stopword)* lebih besar dari pada tanpa penghilangan kata tidak penting *(stopword)*. Namun persentasi kesesuaian kata kunci yang dihasilkan dibandingkan dengan kata kunci yang ada menghasilkan persentase kurang dari 50%, baik yang mengalami penghilangan kata tidak penting maupun yang tidak.

Kriteria kesesuaian kata kunci adalah rangkaian dari kata benda atau kata sifat. Berdasarkan hal tersebut dan dilihat kesesuaian judul dan isi abstrak, maka diperoleh dengan persentase 64,6% dari seluruh data abstrak yang diujikan.

Waktu rata-rata yang diperoleh dalam pengujian pada proses tanta penghilangan kata tidak penting *(stopword)* adalah 0,945692 detik, sedangkan waktu rata-rata yang dibutuhkan adanya proses penghilangan kata tidak penting *(stopword)* adalah 0,591538 detik. Penghilangan kata tidak penting *(stopword)* dapat mempercepat proses eksekusi. Waktu yang dibutuhkan peneliti dan pegawai di lingkungan Pusat Penelitian dan Pengembangan Sumber Daya Air untuk mendapatkan suatu kata kunci sekitar 3-10 menit.