

NutchPropertiesCompleteList

List of Nutch Properties

This list of Nutch configuration properties is intended for development. It includes deprecated properties and properties used only "internally". The list is generated from [nutch-default.xml](#) and Java sources.

Legend:

Def: defined in nutch-default.xml

Used: read or set from Java code

Temp: temporarily used to pass settings (eg, from command-line arguments) to map or reduce jobs

Depr.: deprecated

(owr.): some properties are defined in nutch-default.xml (and may be set in nutch-site.xml) but are overwritten programmatically (tests and benchmarks are excluded), eg. via a command-line argument in some Nutch tools

Property	1.X (master Branch)				2.x (deprecated codebase)			
	Def.	Used	Temp.	Depr.	Def.	Used	Temp.	Depr.
anchorIndexingFilter.deduplicate	X	X			X	X		
any23.content_types	X	X						
any23.extractors	X	X						
arc.url.version	-	X						
batch.proxy.port					-	X		
content.server.port	-	X			-	X		
cosine.goldstandard.file	X	X						
crawl.datum.processor.overdue.time.limit	-	X						
crawl.gen.delay	X	X			X	X		
crawldb.inject.filter.normalize.all	-	X						
crawldb.url.filters	X	X	(owr.)				X	
crawldb.url.normalizers	X	X	(owr.)					
crawldb.url.normalizers.scope	-	X						
creativecommons.exclude.unlicensed	-	X			-	X		
db.fetch.interval.default	X	X			X	X		
db.fetch.interval.max	X	X			X	X		
db.fetch.retry.max	X	X			X	X		
db.fetch.schedule.adaptive.dec_rate	X	X			X	X		
db.fetch.schedule.adaptive.inc_rate	X	X			X	X		
db.fetch.schedule.adaptive.max_interval	X	X			X	X		
db.fetch.schedule.adaptive.min_interval	X	X			X	X		
db.fetch.schedule.adaptive.sync_delta	X	X			X	X		
db.fetch.schedule.adaptive.sync_delta_rate	X	-			X	-		
db.fetch.schedule.class	X	X			X	X		
db.fetch.schedule.mime.file	X	X						
db.ignore.also.redirects	X	X						
db.ignore.external.exemptions.file	X	-						
db.ignore.external.links	X	X			X	X		
db.ignore.external.links.mode	X	X						
db.ignore.internal.links	X	X			X	-		
db.injector.override	X	X	(owr.)					
db.injector.update	X	X	(owr.)					
db.max.anchor.length					X	-		
db.max.outlink.length	X	X						
db.max.outlinks.per.page	X	X			X	X		
db.parsemeta.to.crawldb	X	X			X	-		
db.preserve.backup	X	X						
db.reader.stats.sort	-	X	X		-	X	X	
db.reader.topn	-	X	X					

db.reader.topn.min	-	X	X					
db.score.count.filtered	X	X			X	X		
db.score.injected	X	X			X	X		
db.score.link.external	X	X			X	X		
db.score.link.internal	X	X			X	X		
db.signature.class	X	X			X	X		
db.signature.text_profile.min_token_len	X	-			X	-		
db.signature.text_profile.quant_rate	X	-			X	-		
db.stats.score.quantiles	X	X						
db.update.additions.allowed	X	X			X	X		
db.update.max.inlinks	X	X			X	X		
db.update.purge.404	X	X						
db.update.purge.orphans	X	X						
dc.language	-	X						
domain.statistics.mode	-	X	X		-	X	X	
elastic.cluster					X	-		
elastic.host					X	-		
elastic.index					X	-		
elastic.max.bulk.docs					X	-		
elastic.max.bulk.size					X	-		
elastic.port					X	-		
elasticsearch.conf					-	X		
encodingdetector.charset.min.confidence	X	X			X	X		
exchanges.exchanges.file	X	X						
fail.on.job.failure					-	X		
fetcher.bandwidth.target	X	X						
fetcher.bandwidth.target.check.everyNSecs	X	X						
fetcher.filter.urls	X	X						
fetcher.follow.outlinks.depth	X	X						
fetcher.follow.outlinks.depth.divisor	X	-						
fetcher.follow.outlinks.ignore.external	X	-						
fetcher.follow.outlinks.num.links	X	-						
fetcher.job.resume					-	X		
fetcher.job.sitemap					-	X		
fetcher.job.sitemap.detect					-	X		
fetcher.max.crawl.delay	X	X			X	X		
fetcher.max.exceptions.per.queue	X	-			X	-		
fetcher.maxNum.threads	X	X						
fetcher.min.crawl.delay	X	X						
fetcher.normalize.urls	X	X						
fetcher.parse	X	X			X	X		
fetcher.publisher	X	X						
fetcher.queue.depth.multiplier	X	X			X	X		
fetcher.queue.mode	X	X			X	X		
fetcher.queue.use.host.settings					X	X		
fetcher.redirect.dedupcache.seconds	X	X						
fetcher.redirect.dedupcache.size	X	X						
fetcher.server.delay	X	X			X	X		
fetcher.server.min.delay	X	X			X	X		
fetcher.signature	X	X						
fetcher.store.content	X	X			X	X		
fetcher.store.robotstxt	X	X						
fetcher.threads.fetch	X	X	(owr.)		X	X		
fetcher.threads.per.host.by.ip					-	X		
fetcher.threads.per.queue	X	X			X	X		

fetcher.threads.timeout.divisor	X	X					
fetcher.throughput.threshold.check.after	X	X	(owr.)		X	-	
fetcher.throughput.threshold.pages	X	X			X	-	
fetcher.throughput.threshold.retries	X	X					
fetcher.throughput.threshold.sequence					X	-	
fetcher.timelimit	-	X	X		-	X	X
fetcher.timelimit.mins	X	X			X	X	
fetcher.verbose					X	-	
file.content.ignored	X	-			X	-	
file.content.limit	X	X	(owr.)		X	X	
file.crawl.parent	X	X			X	X	
file.crawl.redirect_noncanonical	X	-			X	-	
free.generator.filter	-	X					
free.generator.normalize	-	X					
ftp.content.limit	X	X			X	X	
ftp.follow.talk	X	X			X	X	
ftp.keep.connection	X	X			X	X	
ftp.password	X	X			X	X	
ftp.server.timeout	X	X			X	X	
ftp.timeout	X	X			X	X	
ftp.username	X	X			X	X	
generate.batch.id					-	X	
generate.count					-	X	
generate.count.mode	X	X			X	X	
generate.curTime	-	X			-	X	
generate.expr	-	X					
generate.fetch.delay.expr	X	X					
generate.filter	-	X			-	X	
generate.hostdb	X	X					
generate.max.count	X	X			X	X	
generate.max.count.expr	X	X					
generate.max.distance					X	X	
generate.max.num.segments	-	X					
generate.min.interval	X	X					
generate.min.score	X	X			X	X	
generate.normalise	-	X			-	X	
generate.partition.seed					-	X	
generate.restrict.status	X	X					
generate.sitemap					-	X	
generate.topN	-	X			-	X	
generate.update.crawldb	X	X			X	X	
gora.buffer.read.limit					X	-	
gora.buffer.write.limit					X	-	
hbase.indexer.commit.size					X	-	
hbase.indexer.mapping.file					X	-	
hbase.indexer.zookeeper.property.clientPort					X	-	
hbase.indexer.zookeeper.quorum					X	-	
headings	X	-					
headings.multivalued	X	X					
hostdb.check.failed	X	X					
hostdb.check.known	X	X					
hostdb.check.new	X	X					
hostdb.concurrency.level					-	X	
hostdb.crawldatum.processors	X	X					
hostdb.dump.field.header	-	X					

hostdb.dump.homepages	-	X						
hostdb.dump.hostnames	-	X						
hostdb.filter.expression	-	X						
hostdb.force.check	X	X						
hostdb.lru.size					-	X		
hostdb.num.resolvers.threads	X	X						
hostdb.numeric.fields	X	X						
hostdb.percentiles	X	X						
hostdb.purge.failed.hosts.threshold	X	X						
hostdb.reading.crawldb	-	X	X					
hostdb.recheck.interval	X	X						
hostdb.string.fields	X	X						
hostdb.url.filter	X	X						
hostdb.url.normalize	X	X						
htmlparsefilter.order	X	X			X	X		
htmlunit.enable.css	X	X						
htmlunit.enable.javascript	X	X						
htmlunit.javascript.timeout	X	X						
http.accept	X	X			X	X		
http.accept.charset	X	X			X	X		
http.accept.language	X	X			X	X		
http.agent.description	X	X			X	X		
http.agent.email	X	X			X	X		
http.agent.host	X	X			X	X		
http.agent.host.cookie.file	X	X						
http.agent.name	X	X	(owr.)		X	X		
http.agent.rotate	X	X			X	X		
http.agent.rotate.file	X	X			X	X		
http.agent.url	X	X			X	X		
http.agent.version	X	X			X	X		
http.auth.file	X	X			X	X		
http.auth.verbose	-	X			-	X		
http.content.limit	X	X	(owr.)		X	X		
http.content.truncated	-	X						
http.content.truncated.reason	-	X						
http.enable.cookie.header	X	X						
http.enable.if.modified.since.header	X	X						
http.log.exceptions.suppress.stack	X	X						
http.max.delays					X	-		
http.partial.truncated	X	X						
http.proxy.exception.list	X	X						
http.proxy.host	X	X			X	X		
http.proxy.password	X	X			X	X		
http.proxy.port	X	X			X	X		
http.proxy.realm	X	X			X	X		
http.proxy.type	X	X						
http.proxy.username	X	X			X	X		
http.redirect.max	X	X						
http.redirect.max.exceeded.skip	X	X						
http.robot.rules.whitelist	X	X						
http.robots.403.allow	X	X			X	X		
http.robots.agents	X	X	(owr.)		X	X		
http.store.responsetime	X	X			X	X		
http.time.limit	X	X						
http.timeout	X	X			X	X		

http.tls.certificates.check	X	X						
http.tls.supported.cipher.suites	-	X			-	X		
http.tls.supported.protocols	-	X			-	X		
http.useHttp11	X	X			X	X		
http.useHttp2	X	X						
http.verbose					X	X		
index.content.md	X	X						
index.db.md	X	X						
index.geoip.licensekey	X	X						
index.geoip.usage	X	X						
index.geoip.userid	X	X						
index.jexl.filter	X	X						
index.links.hosts.only	X	-						
index.links.inlinks.host.ignore	X	-						
index.links.outlinks.host.ignore	X	-						
index.metadata					X	X		
index.metadata.multivalued.fields	-	X						
index.metadata.separator	X	X						
index.parse.md	X	X						
index.replace.regex	X	X						
index.static	X	X						
index.static.fieldsep	X	X						
index.static.keysep	X	X						
index.static.valuesep	X	X						
indexer.add.domain	X	X						
indexer.additional.params	-	X						
indexer.binary.base64	-	X						
indexer.delete	-	X						
indexer.delete.robots.noindex	X	X						
indexer.delete.skipped.by.indexingfilter	X	X						
indexer.indexwriters.file	X	X						
indexer.max.content.length	X	X						
indexer.max.title.length	X	X			X	X		
indexer.nocommit	-	X						
indexer.score.power	X	X			X	X		
indexer.skip.notmodified	X	X						
indexer.url.filters	-	X	X				X	
indexer.url.normalizers	-	X						
indexingfilter.order	X	X			X	X		
injector.current.time	-	X	X		-	X	X	
interactiveselenium.handlers	X	X						
io.file.buffer.size	-	X						
io.serializations	X	-			X	-		
jsoup.extractor.property.file					X	X		
lang.analyze.max.length	X	X			X	-		
lang.extraction.policy	X	X			X	X		
lang.identification.only.certain	X	X			X	X		
lang.index.languages	X	X						
lang.ngram.max.length					X	-		
lang.ngram.min.length					X	-		
libselenium.page.load.delay	-	X						
link.analyze.damping.factor	X	X						
link.analyze.initial.score	X	X						
link.analyze.iteration	-	X	X					
link.analyze.normalize.score	-	X			-	X		

link.analyze.num.iterations	X	X						
link.analyze.rank.one	-	X	X					
link.delete.gone	X	X						
link.ignore.internal.domain	X	X						
link.ignore.internal.host	X	X						
link.ignore.limit.domain	X	X						
link.ignore.limit.page	X	X						
link.score.updater.clear.score	X	X						
linkdb.ignore.external.links	X	X						
linkdb.ignore.internal.links	X	X						
linkdb.max.anchor.length	X	X						
linkdb.max.inlinks	X	X						
linkdb.regex	-	X	X					
linkdb.url.filters	-	X	X				X	
linkdb.url.normalizer	-	X						
linkdb.url.normalizer.scope	-	X						
metatag.description	-	X						
metatag.keyword	-	X						
metatag.keywords	-	X						
metatags.names	X	X			X	X		
mime.type.magic	X	X			X	X		
mime.types.file	X	X			X	X		
mimetype.filter.file	X	X						
moreIndexingFilter.indexMimeTypeParts	X	X			X	X		
moreIndexingFilter.mapMimeType	X	X						
moreIndexingFilter.mapMimeType.field	X	X						
nutch.conf.uuid	-	X			-	X		
nutch.fetch.time	-	X						
org.apache.nutch.webui					-	X		
page.load.delay	X	X						
parse.filter.urls	X	X	(owr.)					
parse.job.force					-	X		
parse.job.resume					-	X		
parse.normalize.urls	X	X	(owr.)					
parse.plugin.file	X	X			X	X		
parse.sitemap					-	X		
parsefilter.naivebayes.trainfile	X	X						
parsefilter.naivebayes.wordlist	X	X						
parsefilter.regex.file	-	X						
parsefilter.regex.rules	-	X						
parser.caching.forbidden.policy	X	X			X	X		
parser.character.encoding.default	X	X			X	X		
parser.html.form.use_action	X	X			X	X		
parser.html.impl	X	X			X	X		
parser.html.line.separators	X	X						
parser.html.outlinks.htmlNode_metadata_name	X	X						
parser.html.outlinks.ignore_tags	X	X			X	X		
parser.html.outlinks.max.target.length					X	X		
parser.skip.truncated	X	X			X	X		
parser.store.text	X	X						
parser.timeout	X	X			X	X		
partition.url.mode	X	X			X	X		
partition.url.seed	-	X	X		-	X		
plugin.auto-activation	X	X			X	X		
plugin.excludes	X	X			X	X		

plugin.folders	X	X			X	X		
plugin.includes	X	X			X	X		
preferred.schema.name							X	
publisher.order	X	-						
rabbitmq.publisher.binding	X	X						
rabbitmq.publisher.binding.arguments	X	X						
rabbitmq.publisher.exchange.name	X	X						
rabbitmq.publisher.exchange.options	X	X						
rabbitmq.publisher.headers.static	X	X						
rabbitmq.publisher.queue.name	X	X						
rabbitmq.publisher.queue.options	X	X						
rabbitmq.publisher.routingkey	X	X						
rabbitmq.publisher.server.uri	X	X						
restapi.auth					X	X		
restapi.auth.ssl.keypass					X	X		
restapi.auth.ssl.storepass					X	X		
restapi.auth.ssl.storepath					X	X		
restapi.auth.users					X	X		
scoring.content.md	X	X						
scoring.db.md	X	X						
scoring.depth.max	X	X						
scoring.filter.order	X	-			X	X		
scoring.orphan.mark.gone.after	X	X						
scoring.orphan.mark.orphan.after	X	X						
scoring.parse.md	X	X						
scoring.similarity.model	X	X						
scoring.similarity.ngrams	X	X						
scoring.similarity.stopword.file	X	X						
screenshot.location	X	X						
segment.dump.dir	-	X						
segment.merger.filter	-	X	X					
segment.merger.normalizer	-	X	X					
segment.merger.segmentName	-	X	X					
segment.merger.slice	-	X	X					
segment.proxy.port	-	X						
segment.reader.content.recode	X	X	(owr.)					
selenium.driver	X	X						
selenium.enable.headless	X	X						
selenium.firefox.allowed.hosts	X	-						
selenium.firefox.binary.timeout	X	-						
selenium.firefox.enable.flash	X	-						
selenium.firefox.load.image	X	-						
selenium.firefox.load.stylesheet	X	-						
selenium.grid.binary	X	X						
selenium.grid.driver	X	X						
selenium.hub.host	X	X						
selenium.hub.path	X	X						
selenium.hub.port	X	X						
selenium.hub.protocol	X	X						
sftp.password					-	X		
sftp.port					-	X		
sftp.server					-	X		
sftp.user					-	X		
sitemap.content.limit					X	-		
sitemap.parser.timeout					X	X		

sitemap.redir.max	X	X						
sitemap.size.max	X	X						
sitemap.strict.parsing	X	X						
sitemap.url.default.sitemap.xml	X	X						
sitemap.url.filter	X	X						
sitemap.url.normalize	X	X						
sitemap.url.overwrite.existing	X	X						
solr.auth					X	X		
solr.auth.password					-	X		
solr.auth.username					-	X		
solr.commit.index					X	X		
solr.commit.size					X	X		
solr.mapping.file					X	X		
solr.server.url					-	X		
storage.crawl.id					X	X		
storage.data.store.class					X	-		
storage.schema.host					X	X		
storage.schema.webpage					X	X		
store.http.headers	X	X						
store.http.request	X	X						
store.ip.address	X	X			X	X		
subcollection.case.insensitive	X	X						
subcollection.default.fieldname	X	X						
subcollection.metadata.source	-	X						
subcollections.config	-	X			-	X		
subcollections.xml	-	X			-	X		
take.screenshot	X	X						
tika.boilerpipe					X	X		
tika.boilerpipe.extractor					X	X		
tika.config.file	X	X						
tika.extractor	X	X						
tika.extractor.boilerpipe.algorithm	X	X						
tika.extractor.boilerpipe.mime.types	X	X						
tika.htmlmapper.classname	X	X			X	X		
tika.parse.embedded	X	X						
tika.uppercase.element.names	X	X						
urlfilter.automaton.file	X	X			X	X		
urlfilter.automaton.rules	-	X			-	X		
urlfilter.domain.file	X	X			X	X		
urlfilter.domain.rules	-	X			-	X		
urlfilter.domainenylist.file	-	X						
urlfilter.domainenylist.rules	-	X						
urlfilter.fast.file	X	X						
urlfilter.order	X	X			X	X		
urlfilter.prefix.file	X	X			X	X		
urlfilter.prefix.rules	-	X			-	X		
urlfilter.regex.file	X	X			X	X		
urlfilter.regex.rules	-	X			-	X		
urlfilter.suffix.file	X	X			X	X		
urlfilter.suffix.rules	-	X			-	X		
urlfilter.tld.length					X	X		
urlmeta.tags	X	X						
urlnormalizer.basic.host.idn	X	-						
urlnormalizer.basic.host.trim-trailing-dot	X	-						
urlnormalizer.hosts.file	-	X						

urlnormalizer.hosts.rules	-	X						
urlnormalizer.loop.count	X	X			X	X		
urlnormalizer.order	X	X			X	X		
urlnormalizer.protocols.file	X	X						
urlnormalizer.protocols.rules	X	X						
urlnormalizer.regex.file	X	X			X	X		
urlnormalizer.regex.rules	-	X			-	X		
urlnormalizer.slashes.file	-	X						
urlnormalizer.slashes.rules	-	X						
warc.exporter.only.successful.responses	-	X						
warc.file.size.max	-	X						
webdriver.chrome.driver	X	-						
webgraph.url.filters	-	X	X				X	
webgraph.url.normalizers	-	X						
webgui.auth.users					X	X		
webtable.dump.content					-	X		
webtable.dump.headers					-	X		
webtable.dump.links					-	X		
webtable.dump.text					-	X		
webtable.url.regex					-	X		

back to [FrontPage](#)