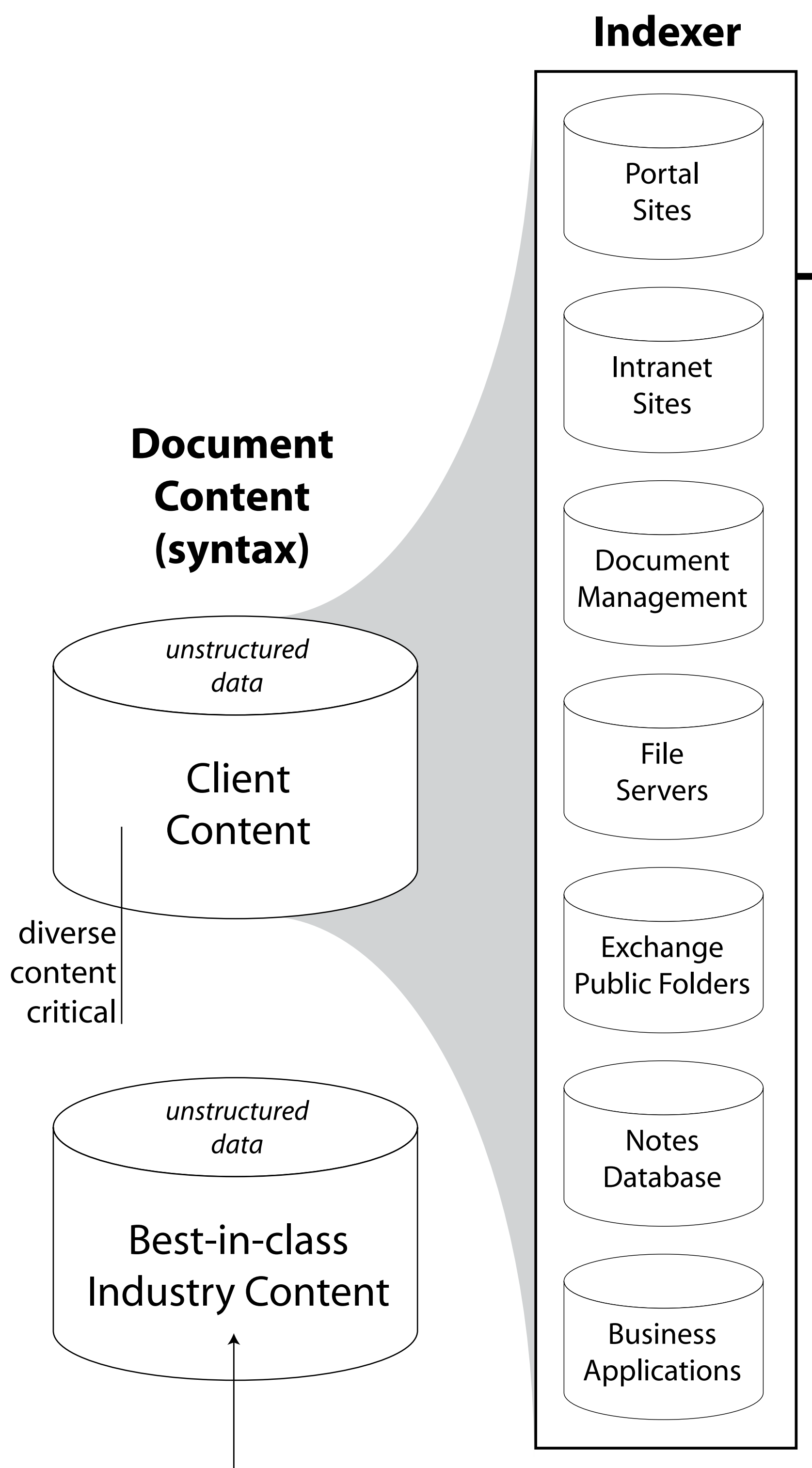
SOPHIE - RULES BUILDER

language sets to organize data



Examples

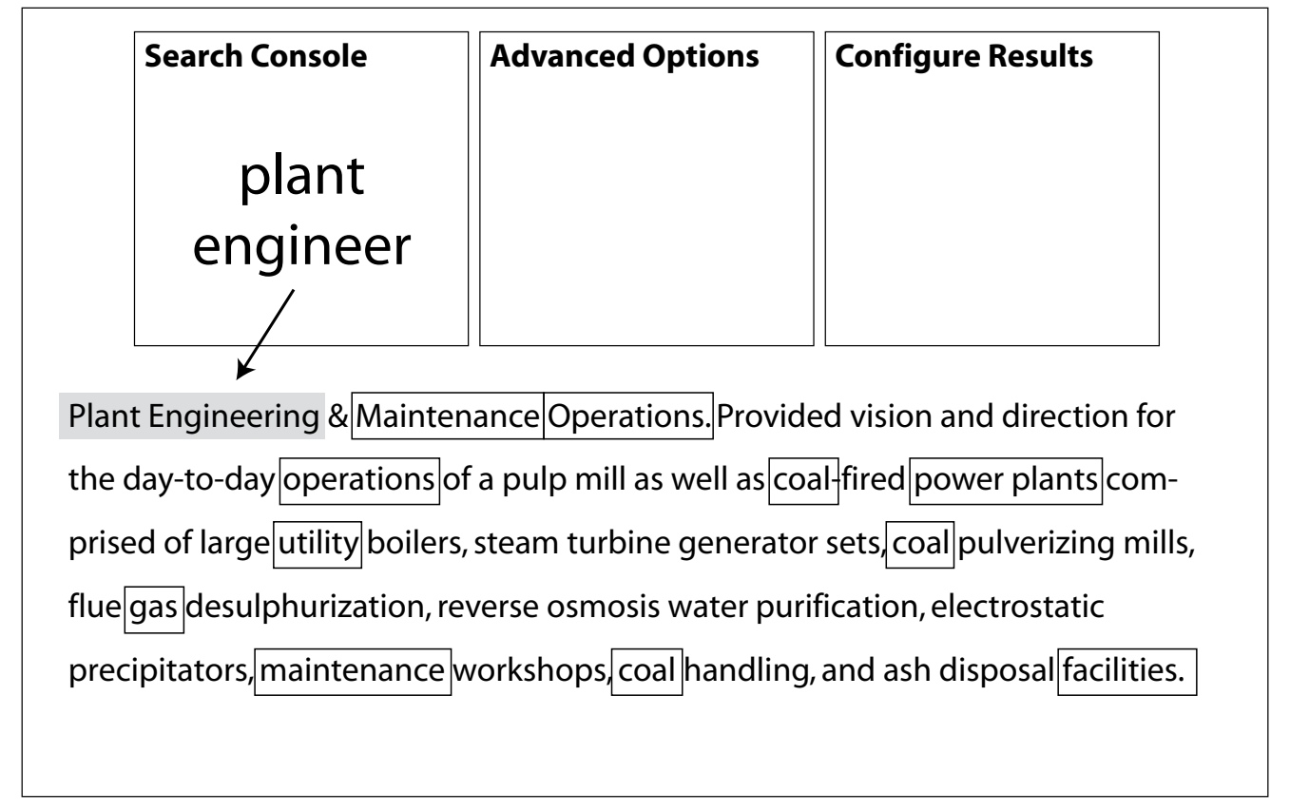
Legal Taxonomy: *Spider sections of US Trademark and Patent Office for taxonomy to describe intellectual property documents at the law firm*

Energy Taxonomy: *Index Oil & Gas Journal sections (E&P, drilling, processing transportation to build best practice language sets*

HC Taxonomy: *Medline article content or other medical web sites that have comprehensive lists of diseases with detailed symptoms and treatments*

Iterative Queries

Start with a simple search....

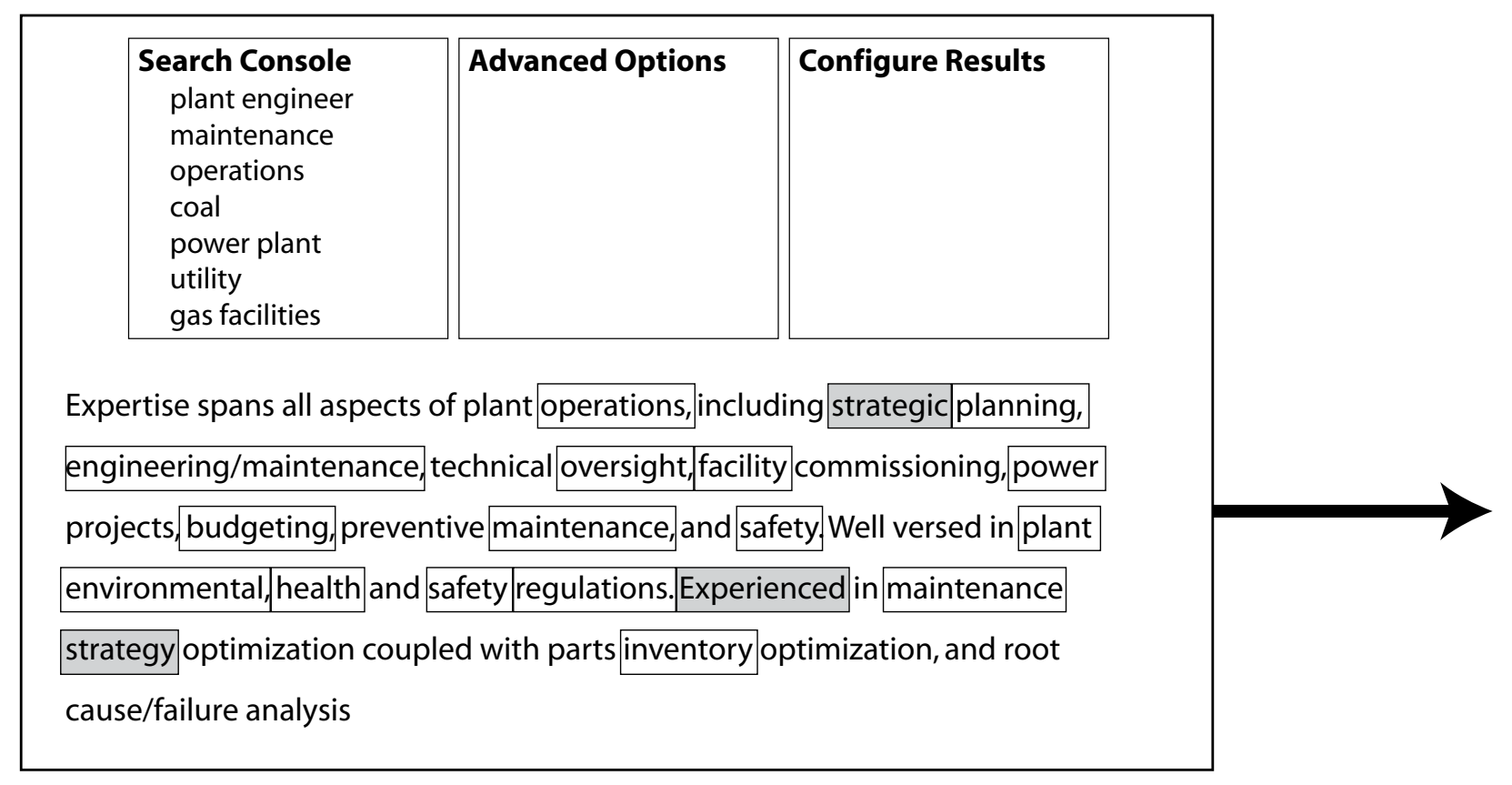


plant engineer

Plant Engineering & Maintenance Operations. Provided vision and direction for the day-to-day operations of a pulp mill as well as coal-fired power plants comprised of large utility boilers, steam turbine generator sets, coal pulverizing mills, flue gas desulphurization, reverse osmosis water purification, electrostatic precipitators, maintenance workshops, coal handling, and ash disposal facilities.

plant engineering = finds the term with focus on phrase and automatic stemming look for nearest neighbor concepts, like maintenance operations, power plants, etc.

Continue with "iterative" search

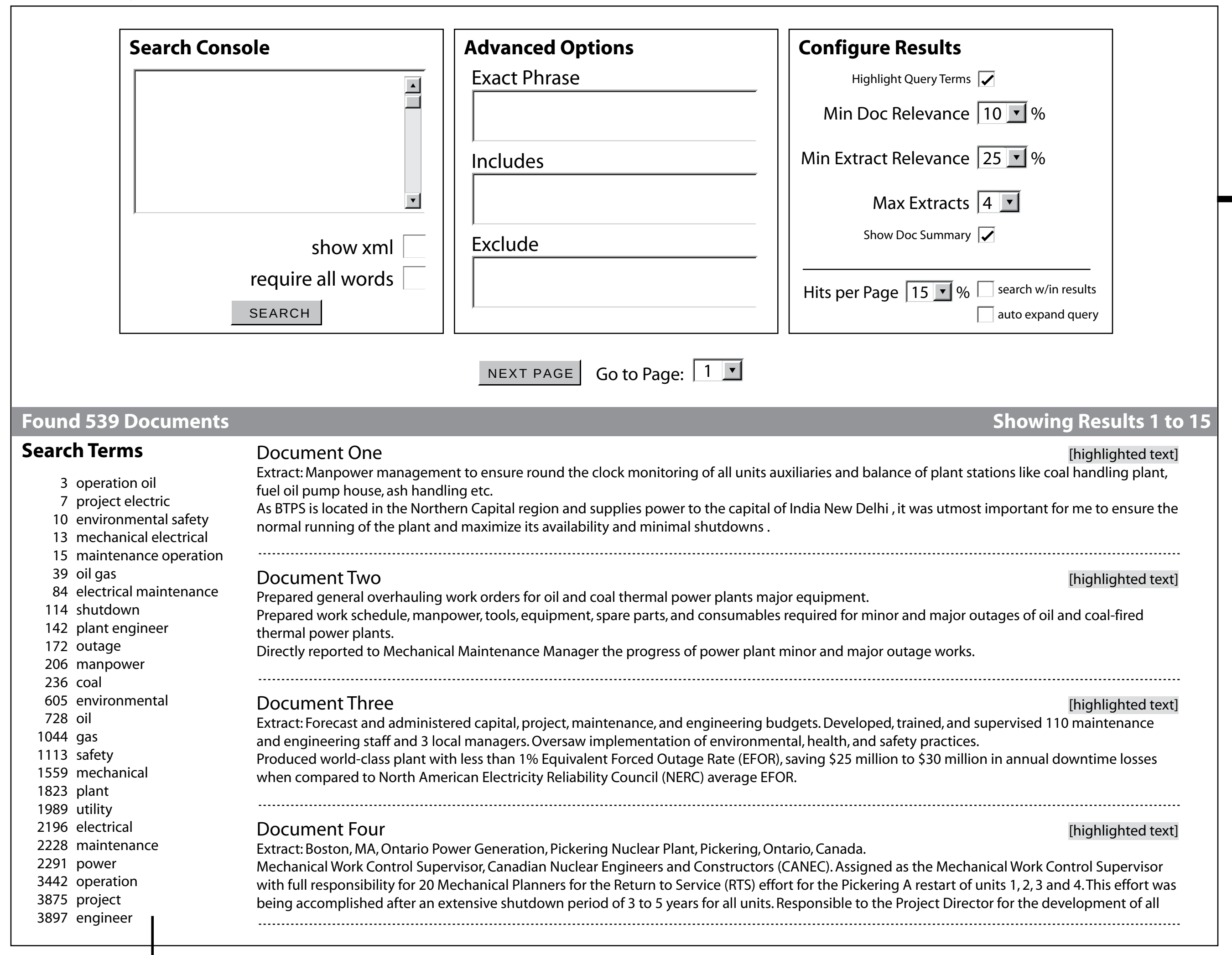


plant engineer
maintenance
operations
coal
power plant
utility
gas facilities

Expertise spans all aspects of plant operations, including strategic planning, engineering/maintenance, technical oversight, facility commissioning, power projects, budgeting, preventive maintenance, and safety. Well versed in plant environmental, health and safety regulations. Experienced in maintenance strategy optimization coupled with parts inventory optimization, and root cause/failure analysis

concepts or keywords — separates concepts — combines keywords into new concepts can search for leadership skills — experience, strategy, etc...

Density Counts & Advanced Functions



Highlight Query Terms ☑
Min Doc Relevance 10 %
Min Extract Relevance 25 %
Max Extracts 4
Show Doc Summary ☑
Hits per Page 15 % □ search w/in results
□ auto expand query

Go to Page: 1

Found 539 Documents — **Showing Results 1 to 15**

Search Terms

3	operation oil
7	project electric
10	environmental safety
13	mechanical electrical
15	maintenance operation
39	oil gas
84	electrical maintenance
114	shutdown
142	plant engineer
172	outage
206	manpower
236	coal
605	environmental
728	oil
1044	gas
1113	safety
1559	mechanical
1823	plant
1989	utility
2196	electrical
2228	maintenance
2291	power
3442	operation
3875	project
3897	engineer

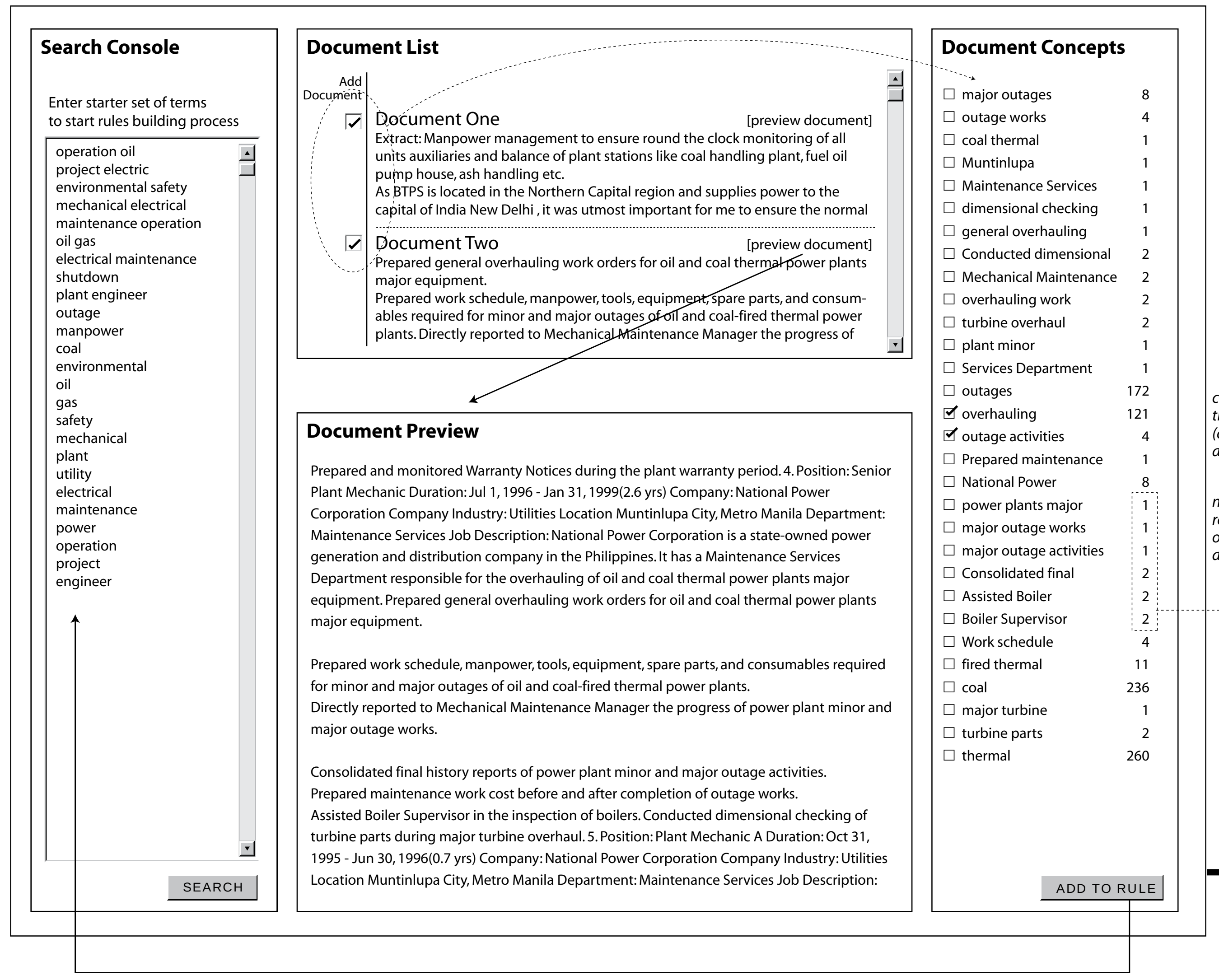
Document One [highlighted text]
Extract: Manpower management to ensure round the clock monitoring of all units auxiliaries and balance of plant stations like coal handling plant, fuel oil pump house, ash handling etc.
As BTPS is located in the Northern Capital region and supplies power to the capital of India New Delhi , it was utmost important for me to ensure the normal running of the plant and maximize its availability and minimal shutdowns .

Document Two [highlighted text]
Prepared general overhauling work orders for oil and coal thermal power plants major equipment.
Prepared work schedule, manpower, tools, equipment, spare parts, and consumables required for minor and major outages of oil and coal-fired thermal power plants.
Directly reported to Mechanical Maintenance Manager the progress of power plant minor and major outage works.

Document Three [highlighted text]
Extract: Forecast and administered capital, project, maintenance, and engineering budgets. Developed, trained, and supervised 110 maintenance and engineering staff and 3 local managers. Oversaw implementation of environmental, health, and safety practices.
Produced world-class plant with less than 1% Equivalent Forced Outage Rate (EFOR), saving $25 million to $30 million in annual downtime losses when compared to North American Electricity Reliability Council (NERC) average EFOR.

Document Four [highlighted text]
Extract: Boston, MA, Ontario Power Generation, Pickering Nuclear Plant, Pickering, Ontario, Canada.
Mechanical Work Control Supervisor, Canadian Nuclear Engineers and Constructors (CANEC). Assigned as the Mechanical Work Control Supervisor with full responsibility for 20 Mechanical Planners for the Return to Service (RTS) effort for the Pickering A restart of units 1, 2, 3 and 4. This effort was being accomplished after an extensive shutdown period of 3 to 5 years for all units. Responsible to the Project Director for the development of all

Rules Building



Enter starter set of terms to start rules building process

operation oil
project electric
environmental safety
mechanical electrical
maintenance operation
oil gas
electrical maintenance
shutdown
plant engineer
outage
manpower
coal
environmental
oil
gas
safety
mechanical
plant
utility
electrical
maintenance
power
operation
project
engineer

☑ Document One [preview document]
Extract: Manpower management to ensure round the clock monitoring of all units auxiliaries and balance of plant stations like coal handling plant, fuel oil pump house, ash handling etc.
As BTPS is located in the Northern Capital region and supplies power to the capital of India New Delhi , it was utmost important for me to ensure the normal

☑ Document Two [preview document]
Prepared general overhauling work orders for oil and coal thermal power plants major equipment.
Prepared work schedule, manpower, tools, equipment, spare parts, and consumables required for minor and major outages of oil and coal-fired thermal power plants. Directly reported to Mechanical Maintenance Manager the progress of

Document Preview

Prepared and monitored Warranty Notices during the plant warranty period. 4. Position: Senior Plant Mechanic Duration: Jul 1, 1996 - Jan 31, 1999(2.6 yrs) Company: National Power Corporation Company Industry: Utilities Location Muntinlupa City, Metro Manila Department: Maintenance Services Job Description: National Power Corporation is a state-owned power generation and distribution company in the Philippines. It has a Maintenance Services Department responsible for the overhauling of oil and coal thermal power plants major equipment. Prepared general overhauling work orders for oil and coal thermal power plants major equipment.

Prepared work schedule, manpower, tools, equipment, spare parts, and consumables required for minor and major outages of oil and coal-fired thermal power plants.
Directly reported to Mechanical Maintenance Manager the progress of power plant minor and major outage works.

Consolidated final history reports of power plant minor and major outage activities.
Prepared maintenance work cost before and after completion of outage works.
Assisted Boiler Supervisor in the inspection of boilers. Conducted dimensional checking of turbine parts during major turbine overhaul. 5. Position: Plant Mechanic A Duration: Oct 31, 1995 - Jun 30, 1996(0.7 yrs) Company: National Power Corporation Company Industry: Utilities Location Muntinlupa City, Metro Manila Department: Maintenance Services Job Description:

Document Concepts

□ major outages	8
□ outage works	4
□ coal thermal	1
□ Muntinlupa	1
□ Maintenance Services	1
□ dimensional checking	1
□ general overhauling	1
□ Conducted dimensional	2
□ Mechanical Maintenance	2
□ overhauling work	2
□ turbine overhaul	2
□ plant minor	1
□ Services Department	1
□ outages	172
☑ overhauling	121
☑ outage activities	4
□ Prepared maintenance	1
□ National Power	8
□ power plants major	1
□ major outage works	1
□ major outage activities	1
□ Consolidated final	2
□ Assisted Boiler	2
□ Boiler Supervisor	2
□ Work schedule	4
□ fired thermal	11
□ coal	236
□ major turbine	1
□ turbine parts	2
□ thermal	260

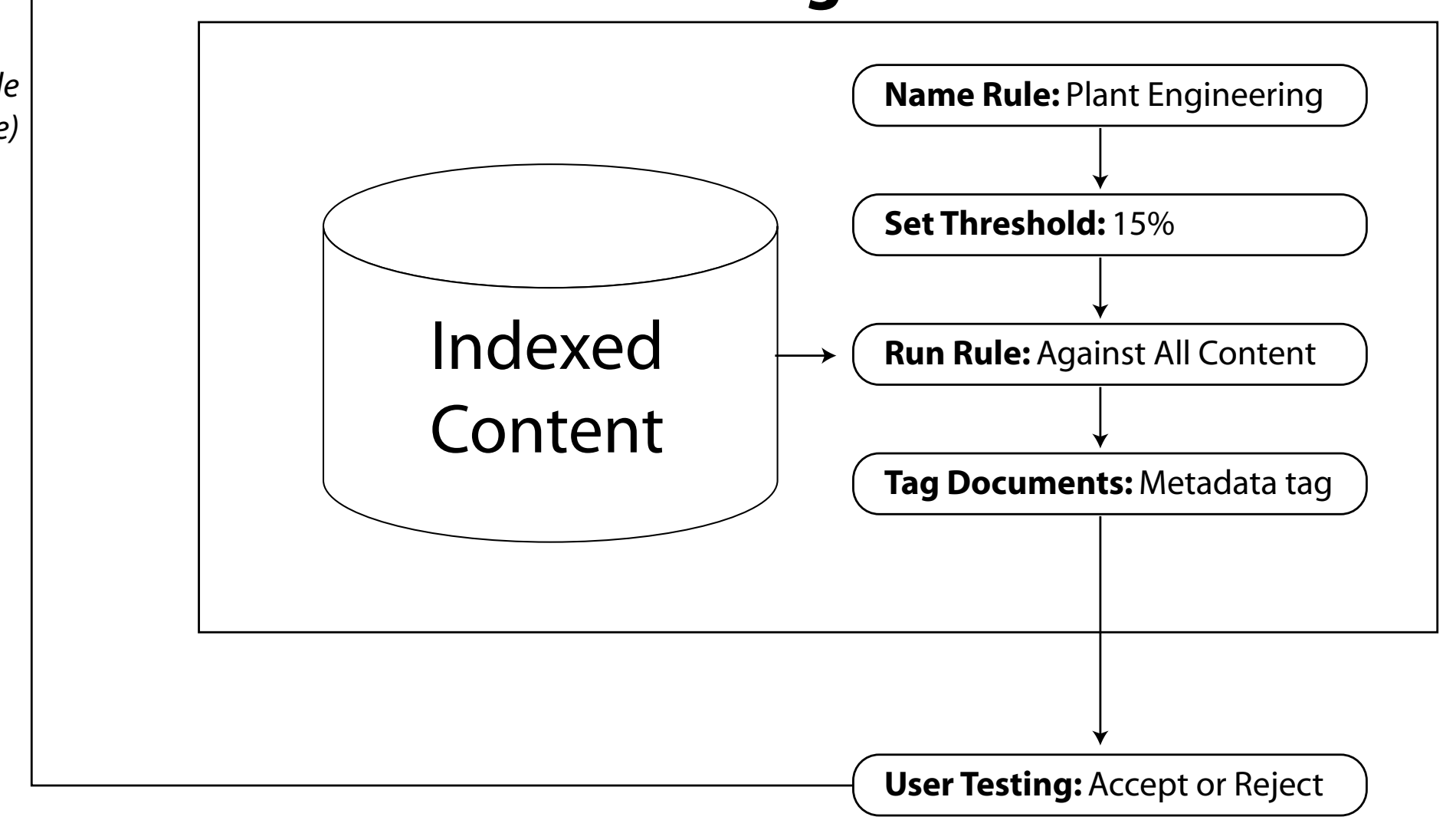
check off concepts from the two checked docs (center pane) and add to the rule

note: many concepts will reduce the relevancy and only pick up 1-2 additional documents

adds overhauling and outage activities to the search query

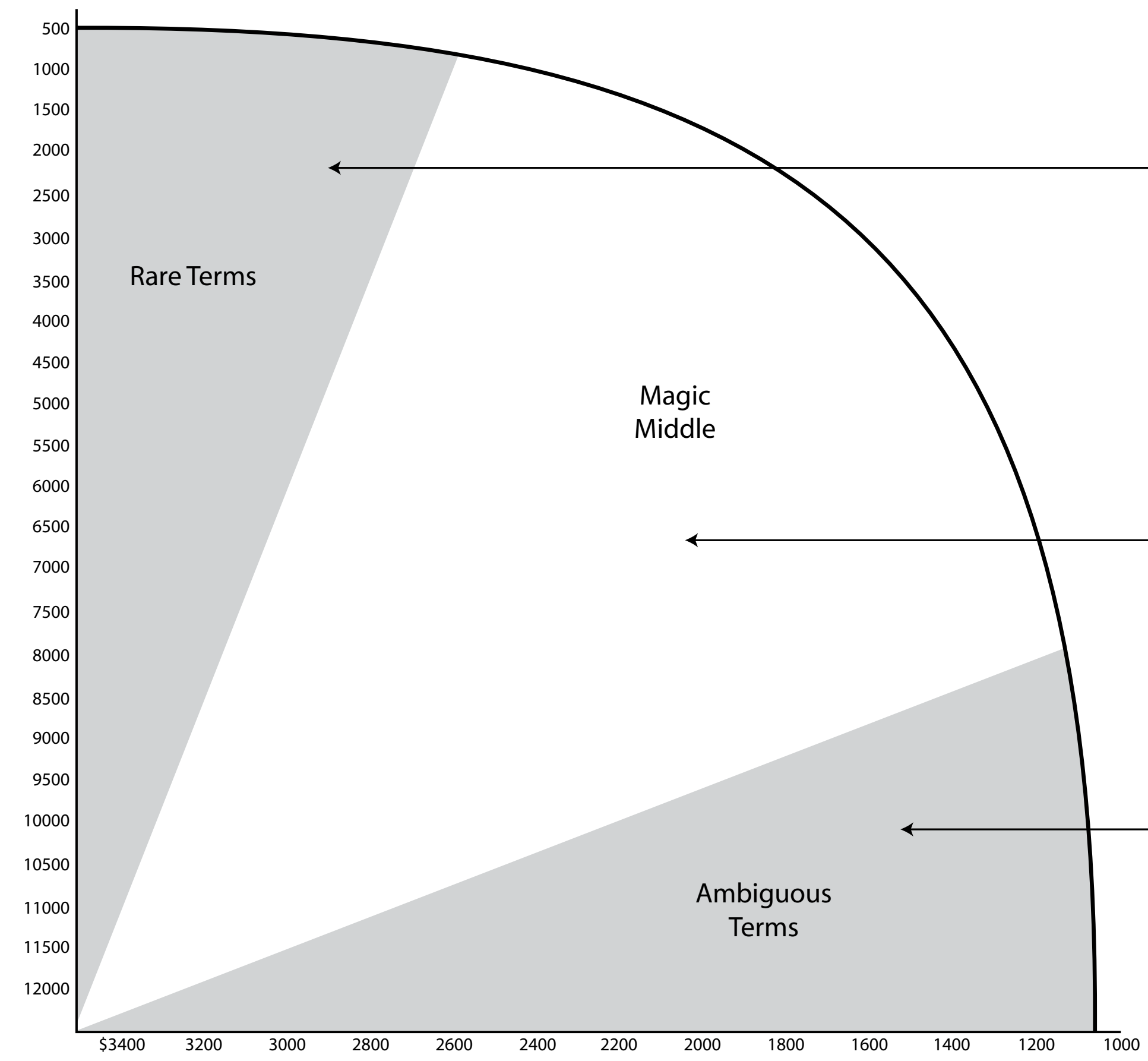
Classification Testing

revise rule (as appropriate)



Indexed Content
Name Rule: Plant Engineering
Set Threshold: 15%
Run Rule: Against All Content
Tag Documents: Metadata tag
User Testing: Accept or Reject

Magic Middle



3	operation oil
7	project electric
10	environmental safety
13	mechanical electrical
15	maintenance operation
39	oil gas
84	electrical maintenance
114	shutdown
142	plant engineer
172	outage
206	manpower
236	coal
605	environmental
728	oil
1044	gas
1113	safety
1559	mechanical
1823	plant
1989	utility
2196	electrical
2228	maintenance
2291	power
3442	operation
3875	project
3897	engineer

some concepts/keywords are so rare, they add no value, but significantly reduce relevance scores

some concepts/keywords are so ambiguous, they appear in most results and are not useful to sequencing relevant results

often very obvious words can hurt the overall query

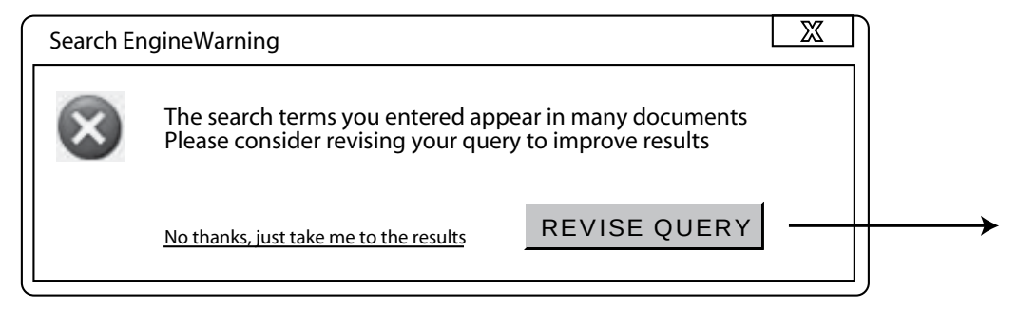
(Weighted) Inverse Document Frequency

Challenge: Does terms in the search query all "weigh" the same as they are submitted? If every document in the collection has the word "oil," then this search term should be meaningless.....

Search String
plant operations SEARCH

plant may appear in 70% of the company's documents operations may appear in 50% of the company's documents therefore, this search is too ambiguous to return good results

coach user to consider revising the query before viewing any results



Search EngineWarning

The search terms you entered appear in many documents
Please consider revising your query to improve results

Search String
plant inventory management SEARCH

much better query string and the user added a third word (mark of improvement)

Challenge: How does the search engine know when search terms (plant operations) will not return meaningful results?

	d_1	d_2	d_3	d_4	d_5
$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$
w_x	2	50	3	2	4
w_y	3	2	3	2	3
$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$

w_x and w_y appear in every document therefore, an argument could be made that these terms do not differentiate the search results...

But w_x in document d_2 does appear much more frequently than in other documents throughout the collection

$$WIDF\ (w,d) = \frac{TF\,(w,d)}{\sum_{i=1}^{n} TF\,(w,d_i)}$$

term frequency of a specific word in a single document

the sum of all mentions of the word across all documents (in a collection or total docs)

Simple example...

$$\text{WIDF}\ (w_x, d_2) = \frac{50}{2+50+3+2+4} = \frac{50}{61} \approx 0.82$$

$$\text{WIDF}\ (w_y, d_2) = \frac{2}{3+2+3+2+3} = \frac{2}{13} \approx 0.15$$

may be able to use WIDF at the collection to universe level or perhaps the query to collection or universe level

will take testing to determine best approach based on need

$$TFIDF\ (w,d) = TF\ (w,d)\ IDF\ (w)$$

$$= \underbrace{TF\ (w,d)}_{} \underbrace{log(\frac{|D|}{\text{DF}\ (w)})}_{}$$

denotes the total number of the documents

the number of documents that contain the term (w)

term frequency handled by search engine (densities)

log(6000/100)

total documents in collection = 6000

query term	*query count*	*docs with term*	*IDF Score*
plant engineer	142	100	1.778
outage	172		
manpower	206		
coal	236		
environmental	605	500	1.079
oil	728		
gas	1044		
safety	1113	350	1.234
mechanical	1559		
plant	1823	1500	0.602

safety appears 1113 times, but is concentrated in just 350 documents making it very material to the query

while "plant" only appears 1823 times in the index, 1500 of the documents contain the term "plant"

another example: assume you are searching resumes and see the word "present" as in 2004-present for the most recent position... the word present would appear in every single resume in the collection (assumed)... making a single mention in the each resume of the results set a very small fraction of the total number of resumes that contain the word "present" (all resumes) — therefore, word is meaningless to accurate language set building

Search String
plant operations

so plant is an ambiguous term (1500 of 6000 docs)

plant inventory management

adding "inventory" helps improve results, even though plant and management are still ambiguous terms

Inverse Document Frequency (IDF) works better than magic middle when user wants to submit an entire document as the search string

Also works well when refine search is offered — improves document extracts before submitting to engine

Load doc as query string

Created and implemented outage and shutdown plans based on production, maintenance, and inspection histories for pulp and power generation plant and systems. Budget Administration. Led the annual development of maintenance, capital, and engineering budgets of up to $20 million. Power Projects. Optimized and managed power plant capital and improvement projects from concept and design to operations in order to meet or exceed project schedules without compromising project goals and profitability. Plant Commissioning. Coordinated plant commissioning with contractors and managed post commissioning engineering issues, while meeting 100% of plant availability targets. Project organization optimization.

Concepts/Keywords Extracted

maintenance inspection	10
power generation plant	13
production maintenance	22
pulp	32
generation plant	38
shutdown	114
outage	172
power generation	192
capital	550
inspection	790
budget	1252
plant	1823
maintenance	2228
power	2291
generation	2779
production	3427
engineering	3897
....	

six mentions

Concepts Applied

Administered capital and maintenance budgets. Managed all plant maintenance functions including engineering, planning, hydraulics maintenance, shutdown planning, and plant re-commissioning. Developed and implemented maintenance and outage project policies, maintenance philosophy, maintenance contracts, and medium/long term outage plan.

notice that "maintenance" is a prevalent term in the top query result but maintenance is in almost every document and therefore ambiguous Magic Middle may remove it because of high count, but... IDF definitely removes it

Concepts De-valued (or valued)

Expertise in work order planning, backlog management, outage and shutdown planning. Inspect plant and equipment, executing unit outages for inspections. Prepare annual O&M budgets, Capital Projects/Budgets, Outage Planning, CSA Reviews and EHS Compliance. Direct and supervise power engineers assigned to assist in plant operation, repairing power generation and converter equipment

removing maintenance from query string can change search results significantly

DSGN *David Gossett*	EDM LOGIC — RULES BUILDING	DIAGNOSTIC
PROG		
DATA	DESC *Rules builder utilizing iterative search and concept building tools, resulting in a language set that returns the highest recall of relevant results, sequenced based on a pattern match to each language set*	
WEB		
APPR		
CONT		
DATE *15-AUGUST-2007*		
COMP **VERSION 1.0**	SIZE 36/54 · CODE EDM TAX	DRAWING NO EDM.TAX.2.WEB · ARCH David Gossett
DEVELOPMENT PURPOSES ONLY	SCALE	SPEC Blacklines · SHEET 2 OF 2

PRIVACY