iraya
ELASTICDOCS

How fast is fast?

Metrics of Machine Learning Enabled-Processing of High Volume Well Reports for Effective Data Search and Class Aggregation in Elastic Docs

Kim Gunn Maver, Nina Hernandez et al, Iraya Energies

FORCE Symposium
20 September 2018
Stavanger, Norway

www.irayaenergies.com
#machinelearninggeoscience

Wells drilled

Region or country	Wells forecast 2018	Wells drilled 2017	% diff.
North America	**198**	**194**	**2.1**
Canada	5	9	-44.4
Cuba	0	0	...
Mexico	44	38	15.8
U.S. - Alaska	12	11	9.1
U.S. - California	28	17	64.7
U.S. - Gulf of Mexico	109	119	-8.4
Others	0	0	...
South America	**89**	**82**	**8.5**
Argentina	1	0	...
Brazil	60	58	3.4
Chile	0	0	...
Colombia	2	3	-33.3
Ecuador	0	0	...
Peru	0	0	...
Trinidad & Tobago	18	16	12.5
Venezuela	2	2	0.0
Others	6	3	100.0
Western Europe	**337**	**315**	**7.0**
Denmark	5	4	25.0
France	0	0	...
Germany	3	1	200.0
Italy	3	1	200.0
Netherlands	14	9	55.6
Norway	210	205	2.4
United Kingdom	93	88	5.7
Others	9	7	28.6
Eastern Europe/FSU	**98**	**90**	**8.9**
Croatia	2		#DIV/0!
Former Soviet Union	90	86	4.7
Russian Federation	n.a.	n.a.	...
Others	90	86	4.7
Poland	1	1	0.0
Romania	5	3	66.7
Others	0	0	...
Africa	**187**	**148**	**26.4**
Angola	50	44	13.6
Congo	28	24	16.7
Egypt	27	20	35.0
Gabon	3	3	0.0
Libya	6	5	20.0
Nigeria	35	28	25.0
South Africa	3	0	...
Tunisia	1	1	0.0
Others	34	23	47.8
Middle East	**321**	**299**	**7.4**
Iran	n.a.	n.a.	...
Neutral Zone	0	0	...
Oman	1	2	-50.0
Qatar	67	61	9.8
Saudi Arabia	74	73	1.4
Turkey	1	0	...
UAE - Abu Dhabi	159	145	9.7
UAE - Dubai	8	8	0.0
Others	11	10	10.0
Far East/South Asia	**1,094**	**986**	**11.0**
Brunei	29	30	-3.3
China	280	230	21.7
India	101	96	5.2
Indonesia	48	46	4.3
Japan	1	1	0.0
Malaysia	57	49	16.3
Myanmar	2	3	-33.3
Pakistan	0	0	...
Philippines	1	1	0.0
Thailand	545	505	7.9
Vietnam	21	20	5.0
Others	9	5	80.0
South Pacific	**31**	**19**	**63.2**
Australia	27	16	68.8
East Timor	2	2	0.0
New Zealand	1	0	...
Papua New Guinea	1	1	0.0
World Total	**2,355**	**2,133**	**10.4**

*Some countries are estimated.

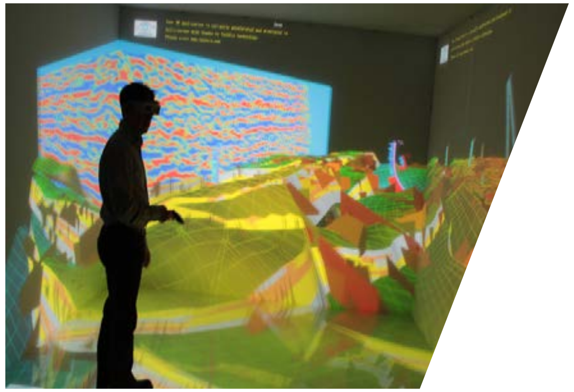
n.a.---Not available.

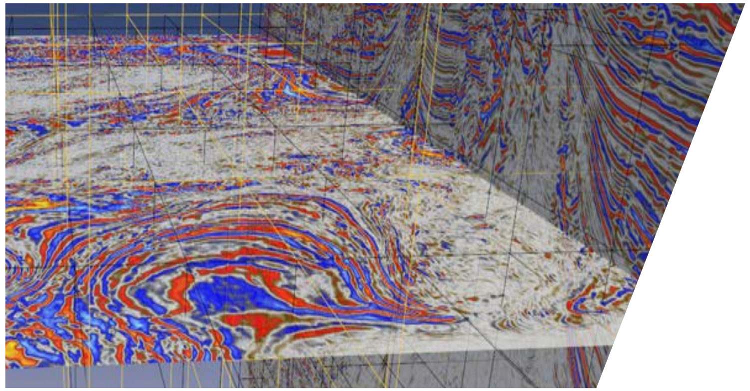
Presentation Outline

- Introduction
- ElasticDocs Workflow
- Machine Learning Metrics
- Experience with a Norwegian Dataset
- Future Developments

Industry challenge

Highly advanced platforms
for well and seismic data





Unstructured
documents
are difficult to access

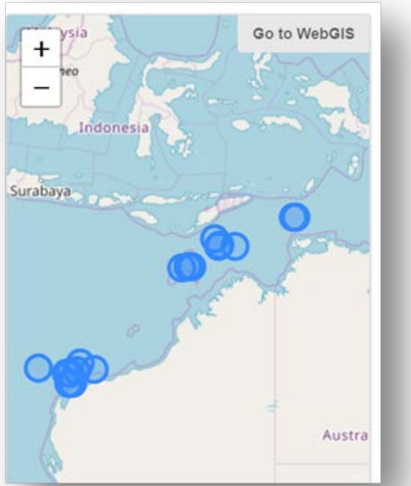
	WELL_NAME	BASIN	PERMIT	OPERATOR	LATITUDE
Abalone1_FWR	Abalone -1	Browse Basin	AC/P29	Operations Summary	[illegible]
Abalone_1_Final_Well_Report_Govt	Abalone—fn	Browse	AC/P 29	J APEX AC Ltdn	[illegible]
Acme 1 FWR	Acme 1	Carnarvon Basin	WA - 205-P\n	Wayne Talbot, Kunal Mishra	[illegible]
Amulet_1_CH1_Well_Completion_Report_Basic	Amulet-1		WA-8-L	Tap Oil Limited	[illegible]
Amulet_2_Well_Completion_Report_Basic_W21078A1	Amulet-2		WA-8-L	Kufpec Australia Pty Ltd	[illegible]
Antiope_1_ST1_Interpretive_Well_Completion_Report_main_text	Antiope — 1/STI	Northern Carnarvon		Antiope-USTI	[illegible]
Appendix_08_Backpacker_1_Bill_Mudlog_FWR	Backpacker-1	Darryl Meakins		Newfield Australia (Cartier) Pty. Ltd.	[illegible]
Argus_1_Well_Completion_Report_Basic_Data_Vol_Main_Text	Argus—1	Ocean General	AC/P30, Argus-1	BHP Petroleum	[illegible]
BALNAVES_DEEP-3_SDL_FWR	Balnaves Deep-3			Apache Energy Limited (AEL)	[illegible]
Backpacker1_FWR	Backpacker-1	Darryl Meakins		Newfield Australia (Cartier) Pty. Ltd.	[illegible]
Balnaves Deep-2_SDL_FWR	Balnaves Deep-2			Apache Energy Limited (AEL)	[illegible]
Banambu Deep-1 Final Well Completion Report	Banambu Deep-1	Analysis Banambu		Banambu Deep-1 Tectonic Elements and Location ...	
Final_Well_Report_MEO_Australia_Artemis_1_FINAL	Artemis-1		WA-360-P	MEO Australia Ltdn	[illegible]
Well Completion Report FINAL	Audacious—5		AC/RL6	OMV Australia Pty Ltd	[illegible]

Decades-worth
of geological insights
LOST

Value Proposition

ElasticDocs Container & Workflow

Information geolocation & density

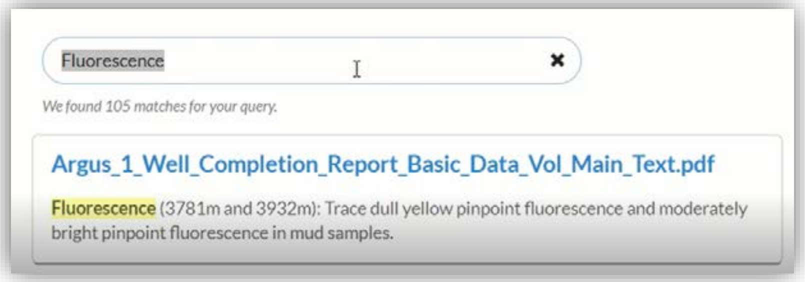


Metadata extraction

	WELL_NAME	BASIN	PERMIT	
Abalone1_FWR	Abalone -1	Browse Basin	AC/P29	Operatio
Abalone_1_Final_Well_Report_Govt	Abalone—lo	Browse	AC/P 29	J AP
Acme 1 FWR	Acme 1	Carnarvon Basin	WA - 205-P\n	Wayne 1
Amulet_1_CH1_Well_Completion_Report_Basic	Amulet-1		WA-8-L	Ta
Amulet_2_Well_Completion_Report_Basic_W21078A1	Amulet-2		WA-8-L	Kufpec Aus
Antiope_1_ST1_Interpretive_Well_Completion_Report_main_text	Antiope — 1/ST1	Northern Carnarvon		A
Appendix_08_Backpacker_1_BH_Mudlog_FWR	Backpacker-1	Darryl Meakins		Newfield Austr
Argus_1_Well_Completion_Report_Basic_Data_Vol_Main_Text	Argus—1	Ocean General	AC/P30, Argus-1	BH
BALNAVES_DEEP-3_SDL_FWR	Balnaves Deep-3			Apache En
Backpacker1_FWR	Backpacker-1	Darryl Meakins		Newfield Austr
Balnaves Deep-2_SDL_FWR	Balnaves Deep-2			Apache En
Banambu Deep-1 Final Well Completion Report	Banambu Deep-1	Analysis Banambu		Banambu Dee Elements an
Final_Well_Report_MEO_Australia_Artemis_1_FINAL	Artemis-1		WA-360-P	MEO Au
Well Completion Report FINAL	Audacious—5		AC/RL6	OMV Aust

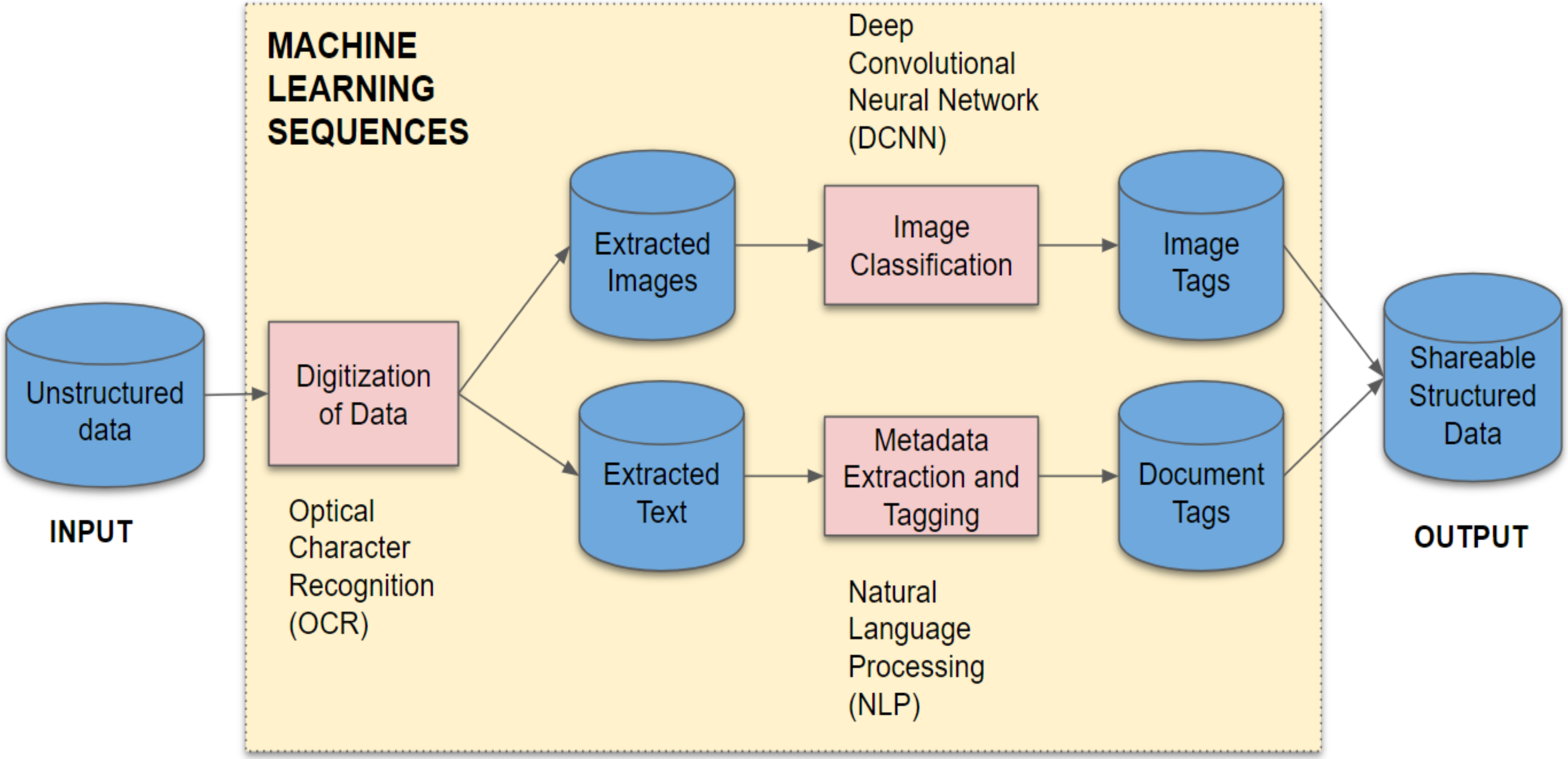
Autoimage clustering & classification

Global elastic search in corpus



ElasticDocs™ Workflow

IMAGE ANALYSIS



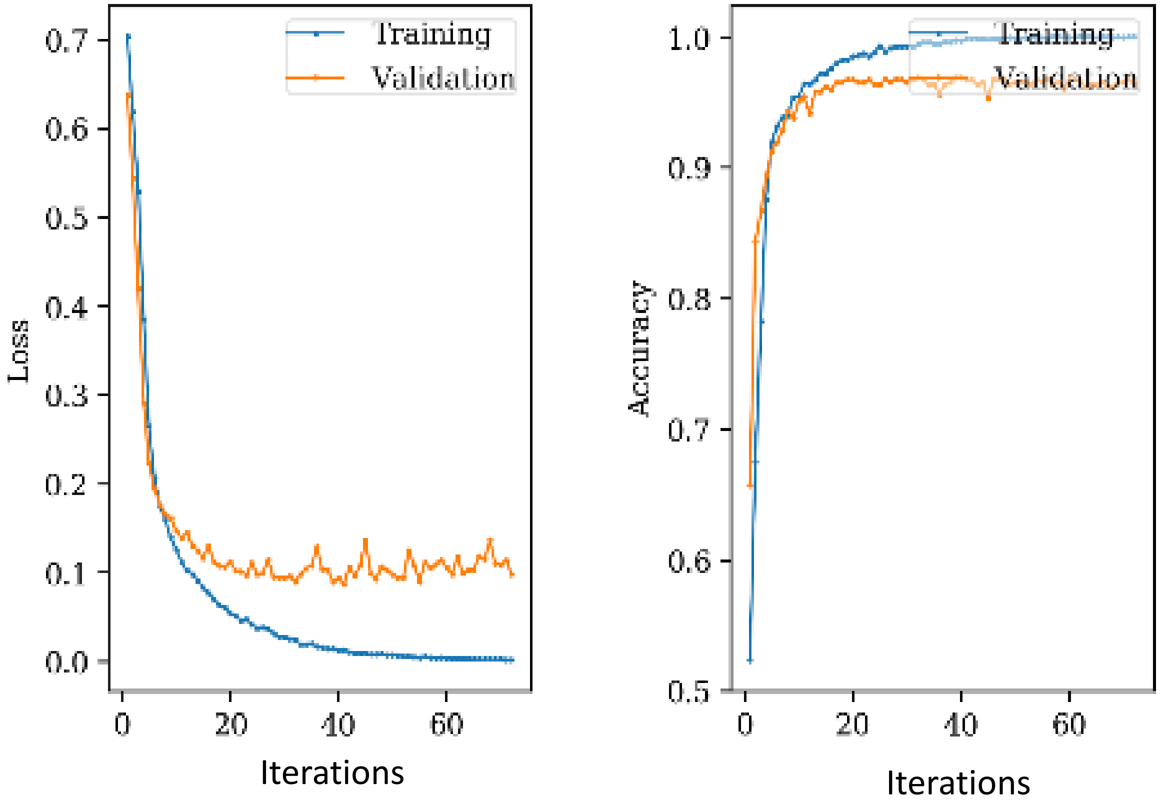
TEXT ANALYSIS

Classification metrics

Machine learning application	Tasks	KPI	Time
Optical character recognition	Text Extraction and Image Extraction	4,542 pages, 6.31 GB, 25 Final Well Reports	10 hours
	Text Extraction only, excluding Image Classification	150,000 pages	3 hours
Deep convolutional neural network	Image Classification	2,598 tagged images input, 16% Tables, 6% Figures, 19% Map, 24% Charts, 33% Noise	20-30 mins during training
Natural language processing	Lithology / Geology Indicator Frequency Analysis (i.e Carbonates, Sandstone, etc)		4 hours
	Well Cataloging	1,500+ input las files, 515 curves identified, 5,681 top log curves (cali, gr, neu, por..)	2,5 hours

Metrics: speed and accuracy

Training Loss



Automated image classification

Supervised image clustering

	Precision	Recall	F1-score
Chart	0.89	0.95	0.92
Core	0.98	1.00	0.99
Figure	0.94	0.72	0.82
Map	0.90	0.98	0.94
SEM	1.00	0.92	0.96
Stratigraphy	0.97	0.85	0.91
Table	0.95	0.96	0.95
Average/total	0.94	0.94	0.94

Precision and Recall

- Precision: proportion of positive identification is correct

$$Precision = \frac{T.P.}{T.P. + F.P.}$$

- Recall: proportion of actual positives is correct

$$Recall = \frac{T.P.}{T.P. + F.N.}$$

- F1 score: harmonic mean of precision and recall

$$F1\ score = \frac{2(precision * recall)}{precision + recall}$$

		Predicted	
		0	1
Actual	0	T.N.	F.P.
	1	F.N.	T.P.

Norwegian dataset

Vintage docs require de-noising filters

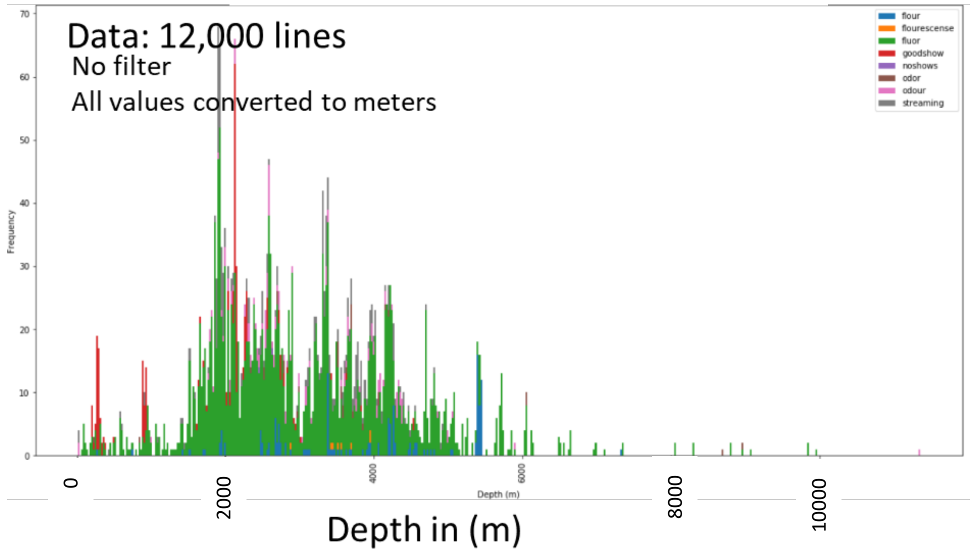
Input:

- 400 pdf files
- 12,000 lines of tagged for hydrocarbon keywords :
  - Flour
  - Fluorescence
  - Fluor
  - Good show
  - No show odor
  - Stain
  - Streaming

Experiment:

- Used Named Entity Recognition to identify depths from the tagged items.

Named entity recognition



Norwegian dataset

Some of the errors in depth values are observed
Original PDF investigated

Reasons:

- Unknown unit of measurements

  `'_ _!n~-~ Fluor fair to good. Cut poor. Gas 300 to 600 units. '`

  Detected: [300,600]

- Error caused by spacing

  `'yel wh res Fluor no vis res. 3 959 as 3958m 3 960 Sst: '`

  Detected: [959,3958,960]

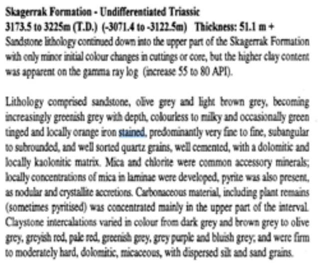
- Error caused by OCR

  `'milky white cut fluor. no cut colour. 332lm - Sandstone: It. gry very fine '`

  Detected: [332]

- Error caused by OCR mode
  Difference in modes for "paragraph" vs "table" format

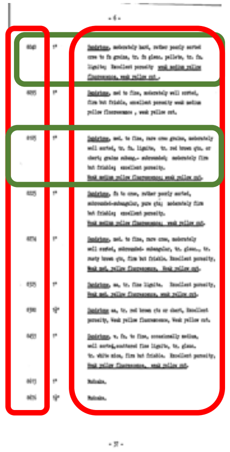
"paragraph"



Skagerrak Formation - Undifferentiated Triassic
3173.5 to 3225m (T.D.) (-3071.4 to -3122.5m) Thickness: 51.1 m +
Sandstone lithology continued down into the upper part of the Skagerrak Formation with only minor initial colour changes in cuttings or core, but the higher clay content was apparent on the gamma ray log (increase 55 to 80 API).

Lithology comprised sandstone, olive grey and light brown grey, becoming increasingly greenish grey with depth, colourless to milky and occasionally green tinged and locally orange iron stained, predominantly very fine to fine, subangular to subrounded, and well sorted quartz grains, well cemented, with a dolomitic and locally kaolonitic matrix. Mica and chlorite were common accessory minerals; locally concentrations of mica in laminae were developed, pyrite was also present, as nodular and crystallite accretions. Carbonaceous material, including plant remains (sometimes pyritised) was concentrated mainly in the upper part of the interval. Claystone intercalations varied in colour from dark grey and brown grey to olive grey, greyish red, pale red, greenish grey, grey purple and bluish grey; and were firm to moderately hard, dolomitic, micaceous, with dispersed silt and sand grains.

"table"



Exciting stuff to add…

- Editable image tags to source documents
- Deeper level of auto-image classifications, i.e:
  - Tables to formation tops
  - Thin sections to mineralogies
- Hydrocarbon show extraction
- Filter by geolocation
- Image search, i.e “google vision”
- Semantic segmentation

Thank you!

Questions?

info@irayaenergies.com
www.irayaenergies.com