

STN[®]

PCTFULL: A New Dimension

Martine MICHEL

CAPADOC

martine.michel@capadoc.fr

Agenda

- PCTFULL Reload
- Numeric Property Search
- Numeric Property Search Examples

PCTFULL Reloaded - providing significant quality enhancements

- Produced by LexisNexis Univentio
- Higher quality of full-text due to improved OCR methods
- Better identification of claims and descriptions
- Improved bibliographic data incl. priority numbers based on EPO-DOCDB
- Updating of backfile records with data additions and data corrections

The new PCTFULL – keeping pace with modern requirements

- Machine translations from several European languages into English
- Display in original Chinese, Korean, Russian
- Searchable address data in IN, PA und AG

PA

POSITEC POWER TOOL (SUZHOU) CO., LTD, **Number 18 Dongwang Road, Industrial Park, Suzhou,**

Jiangsu 215123, CN, [NAT: CN, RES: CN], for all designated states except US;

HUO, Lixiang, Number 18 Dongwang Road, Industrial Park, Suzhou, Jiangsu 215123, CN, [NAT: CN, RES: CN], for US only

Numeric Property Search in PCTFULL

- Numeric Property Data is extracted from the full text and made available for searching by FIZ Karlsruhe for improved retrieval.
 - 34 physical properties in 396 units
 - automatic unit conversion
- Searchable property data is available in all English-language text fields (TIEN, ABEN, CLMEN and DETDEN)
 - proximity searching with numeric properties and keywords
 - flexible data input (open or closed ranges, tolerances, exact values)

Semantic Enrichment: Numbers and Units Behind the Scenes

- The task: make numeric property values and ranges searchable
 - identify numbers and units in the full-text
 - correctly recognize intervals

Semantic Enrichment: Numbers and Units

Identifying Measurements

The resulting CeO_2 particle size measured by x-ray diffraction were in the range of 10 to 30 nm. Fig. 1 shows typical nano particles in a sample milled for 6 hours. In a second experiment a 1 litre attrition mill was used for milling the mixture. ... In addition it is widely accepted that the existence of a so-called 'limiting particle size' limits the practical minimum particle size that can be attained by grinding to values greater than 100nm, irrespective of the type of ball mill employed.

Semantic Enrichment: Numbers and Units

Identifying Measurements

- The task: make numeric property values and ranges searchable
 - identify numbers and units in the full-text
 - correctly recognize intervals
 - extract and normalize to base units

Semantic Enrichment: Numbers and Units Normalisation

values are converted to SI base units

The resulting CeO₂ particle size measured by x-ray diffraction were in the range of 10 to 30 nm. Fig. 1 shows typical nano particles in a sample milled for 6 hours. In a second experiment a 1 litre attrition mill was used for milling the mixture. ... In addition it is widely accepted that the existence of a so-called 'limiting particle size' limits the practical minimum particle size that can be attained by grinding to values greater than 100nm, irrespective of the type of ball mill employed.

Semantic Enrichment: Numbers and Units Normalisation

values are converted to SI base units

10 to 30 nm

1.0×10^{-8} m, 3.0×10^{-8} m

6 hours

2.16×10^4 sec.

1 litre

1.0×10^{-3} m³

greater than 100nm

1.0×10^{-7} m,

Note that for technical reasons, 100nm are included in this range, even though the text defines the size to be 'greater than 100nm'.

Semantic Enrichment: Numbers and Units

Identifying Measurements

- The task: make numeric property values and ranges searchable
 - identify numbers and units in the full-text
 - correctly recognize intervals
 - extract and normalize to base units
 - make this information available for searching

Numeric Property Search in PCTFULL

- The generated numeric properties are searchable in the (English language) full-text
 - combination with text-based search terms
 - standard proximity operators apply
 - searching of exact values and ranges

Numeric Property Search in PCTFULL Units

34 different units are searchable in PCTFULL

Bequerel bit byte candela degree Hertz
Joule Kelvin kg/m^3 kilogram kg/s lumen
Lux m^2 m/s meter m^2/s m^3 mol mol/l
N/m Ohm Pa·s Pascal percent pH rpm
second Siemens Tesla Volt Watt

Numeric Property Search Fields

Property	Field Code	Base Unit
AREA (SURFACE AREA)	/SAR	m ² (square meter)
BIT RATE	/BIR	Bit
BYTE RATE	/BYR	Byte
MOLAR CONCENTRATION (MOLARITY)	/CMOL	mol/L (mol per liter)
CONDUCTANCE (ELECTRICAL CONDUCTANCE)	/CON	S (Siemens)
DEGREE	/DEG	Degree
DENSITY (MASS DENSITY)	/DEN	kg/m ³ (kilogram per cubic meter)
VISCOSITY, DYNAMIC	/DV	PA s (pascal second)
ELECTRICAL IMPEDANCE/RESISTANCE	/RES	ohm
ENERGY	/ENE	J (Joule)
FORCE	/FOR	N (Newton)
FREQUENCY	/FRE	Hz (Hertz)

Numeric Property Search Fields

Property	Field Code	Base Unit
VISCOSITY, KINEMATIC	/KV	m ² /s (square meter per second)
LUMINOUS EMITTANCE/ILLUMINANCE	/LUME	Lux
LUMINOUS FLUX (LUMINOUS POWER)	/LUMF	Lumen
LUMINOUS INTENSITY	/LUMI	Candela
MASS	/M	Kg (kilogram)
MAGNETIC FIELD STRENGTH (MAGNETIC FLUX DENSITY)	/MFS	T (Tesla)
MASS FLOW (MASS TRANSFER)	/MFL	Kg/s (kilogram per second)
MOLECULAR WEIGHT (MOLAR MASS)	/MW	g/mol (gram per mol)
PERCENT (PROPORTIONALITY)	/PER	percent
PH VALUE	/PHV	pH
POWER	/POW	W (Watt)
PRESSURE	/PRES	Pa (Pascal)

Numeric Property Search Fields

Property	Field Code	Base Unit
RADIOACTIVITY	/RAD	Bq (Becquerel)
SPRING CONSTANT	/SCO	N/m (Newton per meter)
SIZE	/SIZ	m (meter)
SURFACE TENSION	/ST	J/m ² (Joule per square Meter)
TEMPERATURE	/TEMP	K (Kelvin)
TIME	/TIM	S (second)
VELOCITY	/VEL	m/s (meter per decond)
VELOCITY, ANGULAR	/VELA	rpm (rotations per minute)
VOLUME	/VOL	m ³ (cubic meter)
VOLTAGE	/VOLT	V (Volt)

For more details about searchable units see:
=>[HELP NPS](#) in PCTFULL

Automatic unit conversion

=> S 30-40 CELSIUS/TEMP

L1 136689 30-40 CELSIUS/TEMP

=> S 30-40 C/TEMP
Is also possible!

=> D KWIC

L1 ANSWER 1 OF 136689 PCTFULL COPYRIGHT 2011 LNU on STN
DETDEN . . .

temperature. The mixture was heated to 35 C for 3 h.. . .

L1 ANSWER 3 OF 136689 PCTFULL COPYRIGHT 2011 LNU on STN
DETDEN . . .

formed hydrogels after incubating at 37 C for 24 hr as. . .

L1 ANSWER 12 OF 136689 COPYRIGHT 2011 LNU on STN
DETDEN . . .

to a temperature from about 50 degrees Fahrenheit to about 550 degrees Fahrenheit.. . .

L1 ANSWER 32 OF 136689 COPYRIGHT 2011 LNU on STN
DETDEN . . .

temperature is within the range 263 Kelvin to 333 Kelvin.

Search example: Nanoscale

```
=> S SIZE (3A) SIZ<=100NM
      756526 SIZE
      112228 SIZ<=100NM
L1    17287 SIZE (3A) SIZ<=100NM
```

```
=> D KWIC 1-5
```

```
L1    ANSWER 1 OF 17287 PCTFULL COPYRIGHT 2011 LNU on STN
```

```
DETDEN . . .
      particles with an average particle size of from approximately 3 to 100
      nanometers, and a. . .
```

```
DETDEN . . .
      30 to 300 or even 100 nanomet
```

```
DETDEN . . .
      SiC powder with an average size of 90 nanometers was added and the. . .
```

```
DETDEN . . .
      TiO2 powder having a particle size of about 50 nanometers is added to
      finely. . .
```

For technical reasons, hit term highlighting will typically only highlight the first number in a range. Units are not highlighted.

Search example: Nanoscale

```
=> S SIZE (3A) SIZ<=100NM
      756526 SIZE
      112228 SIZ<=100NM
L1      17287 SIZE (3A) SIZ<=100NM
```

```
=> S SIZE (3A) SIZ<=1.0E-7
      756526 SIZE
      112228 SIZ<=1.0E-7 M
L2      17287 SIZE (3A) SIZ<=1.0E-7 M
```

Scientific notation is also supported.

```
=> S SIZE (3A) SIZ<=0.0000001
      756526 SIZE
      112228 SIZ<=0.0000001 M
L3      17287 SIZE (3A) SIZ<=0.0000001 M
```

The base unit for size is meter.

Search example: Green AlGaInP LEDs

```
=> S (LIGHT EMITTING DIODE OR LED) (S) AlGaInP (S)  
500-570NM/SIZ
```

```
L1 24 (LIGHT EMITTING DIODE OR LED) (S) ALGAINP (S)  
500-570NM/SIZ
```

Proximities work just as with regular text.

```
=> D KWIC
```

```
L1 ANSWER 3 OF 24 PCTFULL COPYRIGHT 2011 LNU on STN  
DETDEN
```

... The **LED** can be based on different materials, such as, without limitation, GaN, AlGaIn, ..., AlInGaIn (emitting from 285 nm to **550nm**), GaP, GaP:N, GaAsP, GaAsP:N, **AlGaInP** (emitting from **550nm** to 660nm) SiC, ..., (emitting in near infrared and infrared). ...

Search Example: Alloys

```
=> S ALLOY (2S) (BISMUTH (1A) 40-60/PERCENT (S) LEAD (1A)  
PERCENT>20(S) TIN (S) CADMIUM)
```

```
L1          17 ALLOY (2S) (BISMUTH (1A) 40 PERCENT - 60 PERCENT  
/PERCENT (S) LEAD (1A) PERCENT>20 PERCENT (S) TIN (S) CADMIUM)
```

Percentages are also searchable.

```
=> D KWIC
```

```
L1          ANSWER 1 OF 17 PCTFULL COPYRIGHT 2011 LNU on STN
```

```
DETDEN . . .
```

temperature curing processes is used to cover the fasteners. One example of such an alloy is a eutectic alloy, made of about 50% bismuth, about 26.7% lead, about 13.3% tin, and about 10% cadmium by weight, and with a melting point of approximately 70 C (158 F). During the curing. . .

```
L1          ANSWER 2 OF 17 PCTFULL COPYRIGHT 2011 LNU on STN
```

```
DETDEN . . .
```

lowered clearly, if a such alloy contains approx. 14% to 60% bismuth, 20% to 30% lead or up to 45% tin or also antimony, cadmium, Indium, zinc, tellurium, mercury or thallium. In particular with initially the

Search Example: Formulations

```
=> S C11D0001-65+NT/IPC AND (?PHOSPHATE? (5A) PERCENT<0.5)
L1 8 C11D0001-65+NT/IPC AND (?PHOSPHATE? (5A) PERCENT<0.5)
```

Percentages are very useful for searching formulations of pharmaceuticals, pesticides or household chemicals.

```
=> D KWIC
```

```
L1 ANSWER 1 OF 8 PCTFULL COPYRIGHT 2011 LNU on STN
CLMEN. . .
```

STDC0076(b) from 1wt% to 25wt% of an non-alkoxylated anionic deterative surfactant; STDC0040(d) from 0wt% to 4wt% zeolite builder; STDC0046(e) from 0wt% to 4wt% phosphate builder; and STDC0040(f) from 0wt% to 10wt% silicate salt.

```
L1 ANSWER 3 OF 8 PCTFULL COPYRIGHT 2011 LNU on STN
DETDEN
```

The alkaline agent to provide a pH (when diluted to 0.25 to 5% by weight phosphates, metal carbonates, and metal hydroxides, particularly alkali or alkaline metal hydroxides such as NaOH; and KOH. The weight percentage. .

Search example: Allowable % for SEQ ID

```
=> S (SEQ ID OR IDENTITY)/CLM(S)PERCENT>80%
```

```
L1      10052 (SEQ ID OR IDENTITY)/CLM(S)PERCENT>80%
```

```
=> D HIT 2
```

Percent>80% can also be searched as:
Percent>80 or Percent>80 percent (with or
without blanc) or per>80

```
L1      ANSWER 2 OF 10052 PCTFULL COPYRIGHT 2011 LNU on STN
```

```
CLMEN
```

9. The fusion protein of claim 8, wherein the immunoglobulin constant domain

is encoded by a nucleotide sequence having at least 80%, 85%, 90%, 95%, 99% or 100% identity to SEQ ID NO:86, SEQ ID NO:87, or SEQ ID NO:90.

Numeric Property Search in PCTFULL Open Ranges

```
=> S VOLT>0.5
```

```
L10      116530 VOLT>0.5 V
```

```
=> D KWIC
```

```
L10      ANSWER 1 OF 116530 PCTFULL COPYRIGHT 2011 LNU on STN
```

```
DETDEN . . .
```

```
the battery may be at 3.7 volts and be able to. . .
```

```
DETDEN . . .
```

```
that may receive power (e.g., +5 volts) from a digital device. . .
```

```
DETDEN . . .
```

```
example, the battery may provide 3.7 volts and the buck/boost power. . .  
step up the voltage to +5 volts.
```

```
L10      ANSWER 2 OF 116530 PCTFULL
```

```
DETDEN . . .
```

```
voltage is typically from about 2,000 to 80,000 volts. The charge. . .
```

```
DETDEN . . .
```

```
Biax tube. A voltage of 9.2 kV was employed. PTFE was. . .
```

Open ranges are truly open – and will thus retrieve irrelevant results.

Numeric Property Search in PCTFULL Open Ranges

```
=> S IRON (2A) PARTICLE (2A) 1E-7<SIZ
```

```
132395 IRON  
198830 PARTICLE
```

Open ranges will often exceed truncation limits.

```
TERM '1E-7 M <SIZ' EXCEEDED TRUNCATION LIMITS - SEARCH ENDED
```

```
=> S IRON (2A) PARTICLE (2A) 1E-7<SIZ<1E-3
```

```
132395 IRON  
198830 PARTICLE  
350796 1E-7 M <SIZ<1E-3 M
```

Restricting the range will allow the search to complete.

```
L11 19 IRON (2A) PARTICLE
```

Technically, this is not a correct hit. In the Numeric Property Search feature, 'greater than' will always be interpreted as 'greater or equal'.

```
=> D KWIC
```

```
L11 ANSWER 1 OF 19 PCTFULL COPYRIGHT 2011 LNU on STN
```

```
DETDEN . . .
```

```
the red and the black iron oxide the particle size is 90 - 100 nm. . .
```

```
L11 ANSWER 5 OF 19 PCTFULL COPYRIGHT 2011 LNU on STN
```

```
CLMEN. . .
```

This range overlaps with the range searched.

```
the preceding claims wherein the iron containing colloidal particle is  
from 5 nm to 1 mm in. . .
```

Numeric Property Search in PCTFULL Exact Values

```
=> S 100/TEMP
L12          2176 100 K /TEMP

=> S 100+-1/TEMP
L13          3252 100+-1 K /TEMP

=> S L2 NOT L1
L14          1076 L2 NOT L1

=> D KWIC

L14  ANSWER 1 OF 1076 PCTFULL
. . .C, and Al, Low temperature (39 K - 298 K) kinetics. . .

L14  ANSWER 2 OF 1076 PCTFULL COPYRIGHT 2011 LNU on STN
. . .to work between -200C and +400C. The used resin is. . .

=> SET TOL TEMP=1
SET COMMAND COMPLETED

=> S 100/TEMP
L15          3252 100 +-1 K /TEMP

=> S L13 NOT L15
L16          0 L13 NOT L15
```

Note:

Currently, the numeric property search feature is optimized for numeric ranges, searching exact values without tolerance could omit relevant documents.

These results should have been in L1 as well.

Use the SET TOL command to automatically turn exact values into ranges. Use SET TOL TEMP=1 PERM to save this setting for your Login ID.

Numeric Property Search in PCTFULL Search Units

=> S 1-5 INCH/SIZ

L17 251985 1-5 INCH/SIZ

Some imperial units can
also be used for searching.

=> D KWIC

L17 ANSWER 1 OF 251985 PCTFULL COPYRIGHT 2011 LNU on STN

DETDEN . . .

surrounding that value. For example, a dimension disclosed as "40 mm" is intended to mean "about 40 mm".

=> S 100 FAHRENHEIT/TEMP

L18 43153 100 FAHRENHEIT +-1 /TEMP

=> D KWIC

L18 ANSWER 1 OF 43153 PCTFULL COPYR

100° F ≈ 37.77°C, this is a correct hit!

DETDEN . . .

typically range from about 5 C to about 90 C. The water to fabric ratio is typically. . .

Numeric Property Search in PCTFULL

Search Units

=> S 37.77 CELSIUS/TEMP

L19 83325 37.77 CELSIUS +-1 /TEMP

=> S L19 NOT L18

L20 40627 L18 NOT L18

=> D KWIC

L20 ANSWER 1 OF 40627 PCTFULL COPYRIGHT

DETDEN . . .

synthesis and release of 5.75 fmol cDNA in 15 minutes at 37 C for MMLV reverse transcriptase, and the production of 5.0 fmol RNA transcript in 20 minutes at 37 C for T7 RNA polymerase.

L20 ANSWER 2 OF 4

DETDEN . . .

in USP apparatus (40 mesh baskets), 100 revolutions per minute (rpm) in pH 5.8 phosphate buffer (0.1 N HCl) 37 C Samples of 5 milliliters (ml) at each time-point, are taken without media replacement at 1, 2, 4, 6, .

There are considerably more hits for 37.77°C±1 than for 100°F±1. Surely something is wrong?

The interval 37.77°C±1 is relatively bigger than 100°F±1.

37° C ≈ 98.6°F, just outside 100°F±1

Numeric Property Search in PCTFULL

Search Units

```
=> SET TOL TEMP=1%  
SET COMMAND COMPLETED
```

The use of percent-based tolerances can compensate for this.

```
=> S 100 FAHRENHEIT/TEMP
```

```
L21      43153 100 FAHRENHEIT +-1% /TEMP
```

```
=> S 37.77 CELSIUS/TEMP
```

```
L22      43096 37.77 CELSIUS +-1% /TEMP
```

```
=> S L21 NOT L22
```

```
L23      57 L21 NOT L22
```

```
=> D KWIC
```

```
L23      ANSWER 1 37.7°C = 99.14°F (within 100±1%), but  
          37,77°C±1% equals 37,3723 – 38,1477°C, again, this is correct.
```

```
DETDEN . . .
```

```
pH of 2.8, grit, (measured on a 40 µ mesh), of 0.031 % and a Tg, (onset,  
by DSC), of 37.3 C.
```

Numeric Property Search in PCTFULL

Search Options

=> S 50/VOL

L1 454 50 M**3 /VOL

searching with *default units*

=> S 50-60/VOL

L1 599 50 M**3 - 60 M**3 /VOL

searching with *intervals*

=> S 10-30 ML/VOL

L1 114646 10-30 ML/VOL

searching with *other units*

=> S SIZ < 5 MM

L1 463893 SIZ < 5 MM

searching with *open ranges*

=> S 5 MM +-1/SIZ

L1 160136 5 MM +-1/SIZ

searching with *tolerances*

=> S 5 MM +-5%/SIZ

L15 113784 5 MM +- 5%/SIZ

searching with *tolerances in %*

The New PCTFULL Summary

- Higher quality of full-text
 - Better identification of claims and descriptions
 - Improved bibliographic data incl. priority numbers
 - More machine translations into English
 - Searchable address data [=> HELP RLO](#)
- Numeric property searching in all English text fields [=> HELP NPS](#)
 - 34 chemical and physical properties
 - Automatic unit conversion
 - Flexibility in data input (exact values, intervals, tolerances, units, orders of magnitude)

Promotional offer

- In May and June you have the opportunity to test the **new numeric property search feature** in the patent full-text database **PCTFULL** and get a discount of up to **100 EUR** (100 USD).
- Only LoginID holders of the STN Karlsruhe Service Center are eligible for this discount.

How it works:

- Search for numeric properties in PCTFULL.
- Fill in the PCTFULL questionnaire and share your experience with us (http://www.stn-international.de/pctfull_quest.html)
- We will provide you with a credit memo
- Deduct the credit amount from your next STN invoice.

Thank you for your attention

STN[®]

For more information

www.capadoc.com

www.stn-international.com

www.cas.org