Numeric property searching in DWPI\textsuperscript{SM} on STN\textsuperscript{®}

Robert Austin – FIZ Karlsruhe
Agenda

• **New!** Numeric property search in DWPI
• Numeric data identification and normalization
• Numeric property searching
  – Specific property searches
  – Percent searches
  – General searches for records containing properties
Numeric property search in DWPI

- Numeric property searching is now available in Derwent World Patents Index® (DWPI)
  - 1,800 original unit variants indexed
  - 55 numeric property search fields
- Available within all DWPI English-language text fields, e.g. enhanced title, abstract and claims
  - In files WPINDEX, WPIDS and WPIX

NEW!

Type HELP NPS in DWPI on STN to learn more.
Agenda

• Numeric property search in DWPI
• Numeric data identification and normalization
• Numeric property searching
  – Specific property searches
  – Percent searches
  – General searches for records containing properties
Numeric data identification and normalization

• Numbers and their units are identified within the text and made fully numerically searchable
  – 1,800 chemical and physical property unit variants are identified, normalized and indexed
  – Both exact values and ranges

• Identified original data are normalized to base units and indexed for searching
  – 55 numeric property search fields are available

DWPI Numeric property search fields and base units: http://www.stn-international.com/dwpi_nps.html
Example: Numeric data identification and normalization to base units

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 $100$ nm, irrespective of the type of ball mill employed.

Relevant values are converted to SI base units.

Irrelevant numbers are ignored.
Example: Numeric data identification and normalization to base units

Relevant values are converted to SI base units.

- **10 to 30 nm**: $1 \times 10^{-8}$ m, $3 \times 10^{-8}$ m
- **6 hours**: $2.16 \times 10^4$ s
- **1 litre**: $1 \times 10^{-3}$ m$^3$
- **greater than 100nm**: $1 \times 10^{-7}$ m
Agenda

- Numeric property search in DWPI
- Numeric data identification and normalization
- Numeric property searching
  - Specific property searches
  - Percent searches
  - General searches for records containing properties
 Numeric property data can be searched within all English-language text fields

• Numeric search terms can be combined with text-based search terms of interest
  – Using standard text-based proximity operators
  – Specifying text fields of interest, e.g. abstract (/AB)

• Flexible data input options are available
  – 55 chemical and physical numeric fields
  – Ranges, exact values and tolerances
  – A wide variety of STN search units

Search unit ➔ Base unit ➔ Original unit
Searching using the STN units system

- All values searched using accepted units are automatically converted for searching
  - E.g. 100 °C → 373.15 K

- **SET UNIT** to change default search units
  - E.g. => SET UNIT TEMP=F DEN=LB/FT**3** PERM
  - HELP SET UNIT for instructions

- **D UNIT <field>** to see the default and current units for an individual property
  - E.g. => D UNIT TEMP
  - D UNIT ALL to see the complete list
**Valid units systems for searching**

<table>
<thead>
<tr>
<th>System</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>CGS</td>
<td>The centimeter-gram-second system</td>
</tr>
<tr>
<td>ENG</td>
<td>Customary U. S. Engineering units</td>
</tr>
<tr>
<td>FPS</td>
<td>The foot-pound-second system</td>
</tr>
<tr>
<td>MKS</td>
<td>The meter-kilogram-second system</td>
</tr>
<tr>
<td>SI</td>
<td>Systeme Internationale (International System), based on the MKS system</td>
</tr>
<tr>
<td>STN</td>
<td>Customary units based on the SI system</td>
</tr>
</tbody>
</table>

**Tip:** Use e.g. `SET UNITS ALL=CGS` to convert all units to the centimeter-gram-second system.
Numeric operators

- within a range

> greater than

< less than

>= greater or equal to

<= less or equal to
Search example: Particle size

Here we are using the length (/LEN) field* to search for nanoscale particle size (1-100 nm).

Proximity operators may be used just like in a typical text search.

KWIC is a free-of-charge review format for text searches conducted in DWPI subscriber files WPIDS and WPIX.

* Size (/SIZ) may also be used as a synonym for Length (/LEN).
Search example: Particle size (cont.)

=> FILE WPIX

Flexible search input options.

=> S PARTICLE (3A) SIZE (3A) LEN<=100 NM
L1  35606 PARTICLE (3A) SIZE (3A) LEN<=100 NM

The base unit for size is meter.

=> S PARTICLE (3A) SIZE (3A) LEN<=0.0000001
L2  35606 PARTICLE (3A) SIZE (3A) LEN<=0.0000001 M

Scientific notation is supported.

=> S PARTICLE (3A) SIZE (3A) LEN<=1.0E-7
L3  35606 PARTICLE (3A) SIZE (3A) LEN<=1.0E-7 M
Search example: Green AlGaInP LEDs

Here we are using the size (/LEN) field to search for green light wavelengths (500-570 nm).

The same numeric field may be searched in various contexts, e.g. size of wavelength, size of particles, etc.
Search example: Block copolymer molecular weight

=> FILE WPIX

Here we are searching for a Molecular Weight (/MM)* of over 75 kg/mol.

S ?BLOCK? (A) ?POLYMER? /TI,AB (S) MM>75000

L1 52 ((?BLOCK? (A) ?POLYMER?) /TI,AB (S) MM>75000

=> D KWIC 3

Specialized text fields, e.g. the DWPI enhanced title and abstract (/TI,AB), may be used just like in a typical text search.

L1 ANSWER 3 OF 52 WPIX COPYRIGHT 2012 THOMSON REUTERS on STN

NOV NOVELTY - An article comprises a mixture of a polyolefin and a styrenic based block copolymer, where the styrene based block copolymer has an A-B-A block configuration and a molecular weight of at least 350 kg/mol.

Note: in DWPI (S)-proximity limits text terms to the same paragraph.

* /MM = Molar Mass.
Using suitable proximity between numeric and text terms helps find highly relevant hits.
Search example: Dosage

To be more comprehensive, we SELECT the various DWPI Chemistry Resource (DCR) chemical names (/CN) for cisplatin (E1-E40).
PHARMACEUTICALS - Preferred Components: The dosage quantities of each ingredient are 2 mg/kg of cis-platinum, iodized oil, and 0.6 µg/kg of bafilomycin A1 or 60 mg/kg of chloroquine.

Administration - Administration of the cisplatin is 10 mg/kg, intraperitoneally or orally. Administration of the NK1 receptor antagonist is intravenous.

Actives . . . disulfide in physiological saline was administered at 40 mg/kg/day via intraperitoneal (IP) injections. At one hour after MPG disulfide administration, cisplatin (2 mg/kg) was administered to these rats by IP. MPG treatment alone without cisplatin was continued for two more days in . . .
Search example: Unit conversion

=> FILE WPIX

=> S 100000 - 200000/PRES
L1 72205 100000 PA - 200000 PA /PRES

=> S 1 - 2 BAR/PRES
L2 72205 1 - 2 BAR/PRES

=> SET UNIT PRES=BAR
SET COMMAND COMPLETED

=> S 1 - 2/PRES
L3 72205 1 BAR - 2 BAR /PRES

The base unit for pressure (/PRES) is Pascal (Pa).

Values in other accepted search units, e.g. Bar (bar), are automatically converted.

Use SET UNIT to change the default search unit, e.g. from Pascal to Bar.
ABEX. . . at the center of a zirconia-made setter. The compact was thermally treated in an oxygen atmosphere (oxygen partial pressure was 0.1 MPa) at 775 degrees C for 24 hours, to thus produce an independent Li1(Ni(0.75)Co(0.2)Al(0.05))O2 ceramic sheet for a cathode active. . .

T1 is higher than T2 by 5-200 degrees C. The pressure of the interior of the PVT furnace system is 0.1-1500 torr. The method further comprises: doping the ZnO single crystal boule during the PVT growth with at least one positive 100 torr. at 600-1400 degrees C in an oxygen-containing gas environment at an oxygen partial pressure of greater than or equal to 100 torr, and the ZnO wafer have a p-type conduction and a resistivity of less than 0.1 Omega -cm or the. . .
The method involves adopting electric spark cutting technique to cut aluminum sheet with thickness of 1.9 to 2mm. Surface oxide is removed by utilizing corrosion action of deionized water and absolute ethyl alcohol. Adhesive force of...

The method comprises combining polymeric printing raw fabric materials on both sides of aluminum sheet having thickness of 0.25mm then conducting relief and intaglio engraving work over the combined substrate. The obtained material does not need alternative thermal treatment...
An open range voltage (VOLT) search (L1).

Open ranges are truly open – and may retrieve unintended results.

Novelty - The switch has a bipolar transient voltage suppression diode connected to an alternating current (AC) input end of a switch circuit (101). A sensing circuit (103) is sleeved on a coil L (4). The sensing...

Adaptation - Capacitor charging current rectifier device for use with a 12 volt storage battery of an engine. The device is connected with positive and negative pole wires of the 12 volt storage battery to generate electric charge on a coil of a magnetic motor after starting the engine. The device...

Example search: Open ranges
Open range searches may also exceed system limits.

=> S IRON (2A) PARTICLE (2A) LEN>100 NM

TRUNCATION LIMITS EXCEEDED - SEARCH ENDED

=> S IRON (2A) PARTICLE (2A) 100 - 1000 NM/LEN
L2  89 IRON (2A) PARTICLE (2A) 100 - 1000 NM/LEN

=> D KWIC

Restricting the range allows the search to complete.

L2  ANSWER 1 OF 89  WPIX COPYRIGHT 2012       THOMSON REUTERS on STN TECH.

nanoparticle, or their combination. The nanoparticle is an iron oxide magnetic particle. The nanoparticle has an average size diameter of 1-1000 nm. The iron oxide magnetic particle is functionalized through ligand exchange with alpha-carboxyl polystyrene.

Conclusion: it’s typically a good idea to search a closed range, even if you have an open range in mind.
Search example: Exact values

An exact value (L1).

A closed range (L2).

Note: Using exact values (L1) may miss potentially relevant documents (L3).

Conclusion: it’s typically a good idea to search a closed range, even if you have an exact value in mind.
Search example: Tolerances

A tolerance of ± 1 (L1).

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

Percent-based tolerances can also be used (L3, L4).

1% of 100 = 1, so L3 = 99 - 101 °C.

1% of 200 = 2, so L4 = 198 - 202 °C.
Here we are searching for percent (/PER) of metals in an alloy composition.

Inorganic Chemistry - Preferred Components: The metals are lead, zinc, tin, bismuth and cadmium with melting point of 100 degrees C and boiling point of 850 degrees C. The alloy of heat-carrier and/or the metal melt contains bismuth (50%), lead (25%) and tin (25%).
Search example: Monomer weight percent

Here we are searching for the weight % (/PER) of monomers in a copolymer.

=> FILE WPIX

=> S COPOLYM? (S) (30-50/PER (1A) ETHYLENE (NOTW) OXIDE) (5A) (50-70/PER (1A) PROPYLENE (NOTW) OXIDE)

L1 183 COPOLYM? (S) (30 PERCENT - 50 PERCENT /PER (1A) ETHYLENE (NOTW) OXIDE)(5A)(50 PERCENT - 70 PERCENT /PER (1A) PROPYLENE (NOTW) OXIDE)

=> D KWIC

L1 ANSWER 1 OF 183 WPIX COPYRIGHT 2012 THOMSON REUTERS on STN NOV. . . lower than 20 wt.%) comprising a propylene homopolymer and/or a propylene copolymer with at least one alpha-olefin comonomer; copolymer of ethylene (42-70 wt.%) with propylene (35-85 wt.%) (B); polymer composition (5-45 wt.%) (C) comprising a blend of ethylene copolymer (20-80 wt.%) containing at least one alpha-olefin comonomer, and . . .
Here we are searching for the percent (/PER) identity of disclosed sequences to a known protein.

**Search example: Percent sequence identity**

```plaintext
=> FILE WPIX

=> S (FIBROBLAST GROWTH FACTOR OR FGF?) (10A) (SIMILAR? OR IDENTI? OR HOMOLOG?) /TI,AB,TECH(5A) PER>=80

L1 38 (FIBROBLAST GROWTH FACTOR OR FGF?) (10A) (SIMILAR? OR IDENTI? OR HOMOLOG?) /TI,AB,TECH(5A) PER>=80 PERCENT

=> D KWIC

L1 ANSWER 1 OF 38 WPIX COPYRIGHT 2012 THOMSON REUTERS on STN TECH.

bone morphogenetic protein (BMP) polypeptide or its fragment, preferably BMP6 or BMP7 or their polypeptide or fragment. Preferred Sequence: The FGF6 polypeptide exhibits at least 95% homology to a 208 amino acid sequence (SEQ ID NO: 1) fully defined in the specification, and FGF9 exhibits at least 95% homology to a 208 amino acid sequence (SEQ ID NO: 2) fully defined in the specification. The FGF2 polypeptide exhibits at least 95% homology to a 288 amino acid sequence (SEQ ID NO: 3) fully defined in the specification. Preferred Cell: The target cell.
```
in the acidic solution. Reaction temperature of vanadium pentoxide and reducing agent in the acidic solution is 45-100 degrees C. Reacting time is 0.5-64 hours. The pH value of the acidic solution is 0-6. Pressure of reaction in high pressure kettle is 0.1-50 MPa.

Teflon vacuum pump was turned on and the vacuum was immediately applied to the system. After a total reaction time of 60 minutes the heat was turned off and the flask was backfilled with nitrogen. The reaction mixture was then charged with...
reagent is solid sodium methoxide or potassium methoxide or their mixture, the reaction temperature is 70 to 140 degrees centigrade, the reaction time is 1 to 5 hours, the pressure is -0.085 to -0. MPa, removing reaction generated in the methanol.

[CLAIM 5] Preparation method butyl. . .
More focused searches are possible using DWPI abstract sections, e.g. the advantage (/ADV).

ADVANTAGE - The method can greatly enhance catalytic activity, has reaction time of 1.5 hours and quickly form high molecular weight product at maximum molecular weight of more than 20; has body producing, simple...

ADVANTAGE - The reactor system does not use oxygen, has high temperature heating rate (1000-10000 K/s) and has short reaction time (less than 2 seconds). It is capable of providing short chain low molecular substance product and improves yield and quality of biological oil.

See Appendix (slide 40) for a summary of the various DWPI abstract section search options.
Search example: Reaction time (cont.)

=> S REACT? TIME/ADV (5A) **TIM.EX<2 HOURS**
L5  137 REACT? TIME/ADV (5A) **TIM.EX<2 HOURS**

=> S L4 NOT L5
L6  16 L4 NOT L5

Option: exclude indexed open ranges (.EX), as this may help focus the search even further.

Open range hits are not always relevant (L6).

L6 ANSWER ... OF 16 WPIX COPYRIGHT 2012 THOMSON REUTERS on STN
ADV. . . . can be processed for obtaining the Na2Ta2O6 while the hydrothermal reaction temperature is lower than 150 degrees C and the reaction time is **less than 8 hours**.

L6 ANSWER ... OF 16 WPIX COPYRIGHT 2012 THOMSON REUTERS on STN
ADV. . . . greater than 30 (preferably greater than 50) kg. The process of preparing omeprazole form B is carried out at total reaction time of **less than 35 hours**; is simple, cost-effective and large scale applicable; and has improved purification step.
Search example: Physical properties (PHP)

=> FILE WPIX

=> S PER/PHP (5A) CAMPHOR

=> D KWIC 1,20

L1 ANSWER 1 OF 662  WPIX COPYRIGHT 2012       THOMSON REUTERS on STN
NOV. . . . 0.5% bee wax, 6% aluminium starch octenyl succinate, 0.5 % shea butter, 3% triethyl citrate, 0.5 % natural menthol and 0.5 % natural camphor together with usual ingredients and excipients for the remainder to 100%.

L1 ANSWER 20 OF 662  WPIX COPYRIGHT 2012       THOMSON REUTERS on STN
DETD . . . revolutions per minute, cross linking for 5-8 hours to form an uniform microcapsule solution, where the microcapsule emulsifying solution comprises: 1-10% of peppermint camphor, 3-5% of chitosan, 6-8% of modified starch and 0.1-0.5% of emulsifying agent, and (g) spray drying the microcapsule emulsifying solution and emulsifying . . .
labeled substance. The liposome composite complex has surface charge of is -30 to +10 mV. The particle size of the liposome composite complex is 50-500 nm. Preferred Method: The method further involves mixing the liquid with cell transduction substance.

Preferred Liposome: The liposome is of a size 30 - 1500 nm. The acyl chains are myristoyl (14C per chain), or palmitoyl (16C per chain). Stirring. This produces a more fluid and translucent suspension of smaller liposomes. The liposomes encapsulating the oligonucleotide were then extruded. Liposome size was verified to be 112 nm. The solution (1 ml) was then diluted with isotonic saline (2 ml) to a final lipid concentration of 33. . .
Typical numeric searches

Searching for **specific properties**.

=> S (LIGHT EMITTING DIODE OR LED) (10A)
   (AlGaInP or AlInGaP or InGaAlP) (10A) 500–570 NM /LEN

Searching for **percent values**.

=> S ALLOY (S) BISMUTH (1A) 40–60/PER (S) LEAD (1A) PER>20
   (S) TIN (S) CADMIUM

Presence of **numeric properties**.

=> S PER/PHP (5A) CAMPHOR
Review of numeric search options

=> S 50/VOL
L1 4757 50 M**3 /VOL

Searching with default units.

=> S 50-60/VOL
L1 4847 50 M**3 - 60 M**3 /VOL

Searching with closed ranges.

=> S 10-30 ML/VOL
L1 117892 10-30 ML/VOL

Searching with other units.

=> S LEN < 5 MM
L1 1063904 LEN < 5 MM

Searching with open ranges.

=> S 5 MM +-1/LEN
L1 291045 5 MM +-1/LEN

Searching with tolerances.

=> S 5 MM +-5%/LEN
L1 258720 5 MM +- 5%/LEN

Searching with tolerances in %
Summary

- Numeric property search (NPS) is now available in Derwent World Patents Index (DWPI)
- Search for specific units, percentages, or the presence of numeric values within the text
- Combine keyword and numeric terms within the text using standard STN proximity operators
- Search within all DWPI English-language fields, e.g. enhanced title, abstract and claims

DWPI Numeric property search fields and base units: [http://www.stn-international.com/dwpi_nps.html](http://www.stn-international.com/dwpi_nps.html)
Resources

• General information about physical properties given in base or derived SI units
  http://www.bipm.org/en/si

• General Information on the STN Units System
  http://www.cas.org/support/stngen/doc/stnunits/

• DWPI database summary sheet
  http://www.stn-international.com/wpindex.html
# Appendix: DWPI enhanced abstract precision searching options

<table>
<thead>
<tr>
<th>Field</th>
<th>Description</th>
<th>Availability</th>
</tr>
</thead>
<tbody>
<tr>
<td>TI</td>
<td>Enhanced Title</td>
<td>1963-</td>
</tr>
<tr>
<td>NOV</td>
<td>Novelty</td>
<td>1999-</td>
</tr>
<tr>
<td>DETD</td>
<td>Detailed Description</td>
<td>1999-</td>
</tr>
<tr>
<td>ACTN</td>
<td>Mechanism of Action</td>
<td>1999-</td>
</tr>
<tr>
<td>ACTV</td>
<td>Activity</td>
<td>1999-</td>
</tr>
<tr>
<td>ADV</td>
<td>Advantage</td>
<td>1984-</td>
</tr>
<tr>
<td>UADV</td>
<td>Use/Advantage</td>
<td>1984-</td>
</tr>
<tr>
<td>USE</td>
<td>Use</td>
<td>1984-</td>
</tr>
<tr>
<td>DRWD</td>
<td>Drawing Description</td>
<td>1999-</td>
</tr>
<tr>
<td>TECH</td>
<td>Technology Focus</td>
<td>1999-</td>
</tr>
<tr>
<td>ABEX*</td>
<td>Extension Abstract</td>
<td>1999-</td>
</tr>
</tbody>
</table>

* ABEX is only displayable in WPIX.
For more information ...

CAS
E-mail: help@cas.org
Support and Training: www.cas.org

FIZ Karlsruhe
E-mail: helpdesk@fiz-karlsruhe.de
Support and Training: www.stn-international.de

STN is available through FIZ Karlsruhe, Germany and Chemical Abstracts Service, U.S.A.