



When does a search in full-text and chemical databases fail?

Monika Hanelt - Agfa Graphics NV



STAY AHEAD. WITH AGFA GRAPHICS.

Overview

- Introduction - my background
- Failed searches- cases
 - Literature
 - Patents
- Gear-wheels of a search
- Image information
- Outlook

Agfa Graphics in the Agfa-Gevaert Group



GLOBAL TURNOVER 2008
3 billion euro



50%

AGFA GRAPHICS



40%

AGFA HEALTHCARE



10%

AGFA MATERIALS

Agfa Graphics - Market Segments and Applications

- **PREPRESS**

- Commercial Printing
- Packaging
- Newspapers



- **DIGITAL PRINTING**

- Industrial Inkjet
 - Wide Format
 - Medium Web

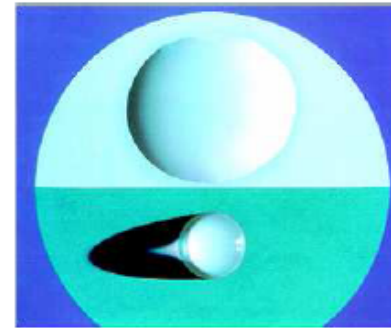


Printing plates - an Application of Chemistry

Application strongly based on Chemistry

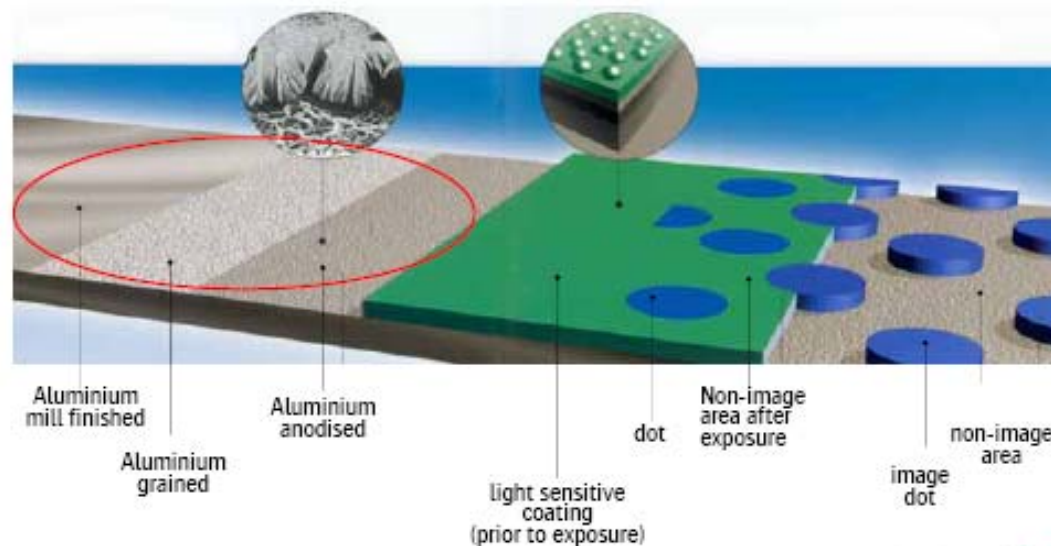
➡ Planographic printing is based on the chemical-physical differences between:

- non-image areas
(hydrophilic = water attracting)
and
- image areas
(oleophilic = grease attracting)



➡ Therefore printing requires:

- a fountain solution (water + additives)
and
- a „greasy“ ink (oil/grease containing)



Typical Patent Applications in our domain

- Typical claims of these application patents:

EP1591242:

1. A lithographic printing plate precursor comprising a photosensitive layer comprising:

- (A) a fluorescent brightening agent;
- (B) an activator compound being capable of inducing a chemical change by an interaction with light absorption of the fluorescent brightening agent to produce at least one of a radical, an acid and a base;
- (C) a compound being capable of undergoing a reaction by an effect of at least one of a radical, an acid and a base to irreversibly change in its physical or chemical properties; and
- (D) a polyurethane resin binder,

wherein the polyurethane resin binder is synthesized from at least following compounds (i), (ii), (iii) and (iv):

- (i) a diisocyanate compound;
- (ii) a diol compound having at least one carboxyl group;
- (iii) a diol compound having a logP value of less than 0; and
- (iv) a diol compound having a logP value of 0 or more,

with the proviso that each of the diol compound (iii) and the diol compound (iv) does not have a carboxyl group.

or

WO2005044937:

1. A non-aqueous, radiation-curable ink comprising pigment, dispersant and a liquid carrier, wherein the pigment has an acid value greater than 8 mg of NaOH per gram of pigment.
2. An ink according to claim 1 having a viscosity of 1 to 30 mPa.s at 60°C.

Our typical Chemical Searches

- Find a chemical compound with a specific structure and specific chemical property
- Which class of surfactants, dispersants, polymers etc has property A in the range of ...
- Specific compounds or compounds classes as such and/or use for imaging applications

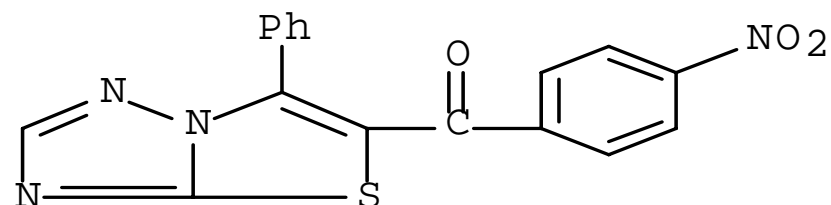
Data Repository for our searches

- Full text databases – not indexed databases
 - Patent full-text databases
 - Literature full-text databases

- Chemical databases - Indexed databases
 - Indexed literature references
 - Indexed patent references

A simple exemplary question

- Find melting point for the following thiazolo-triazole compound:



The search in CAS for the substance delivered the substance and the following table:
no measured data, only predicted data!

Predicted Properties (PPROP)

PROPERTY (CODE)	VALUE	CONDITION	NOTE
Density (DEN)	1.52+/-0.1 g/cm**3	20 deg C 760 Torr	(1)
Freely Rotatable Bonds (FRB)	4		(1)
H acceptors (HAC)	7		(1)
H donors (HD)	0		(1)
Hydrogen Donors/Acceptors Sum (HDAS)	7		(1)
LOGP (LOGP)	4.129+/-1.141	25 deg C	(1)
Molar Volume (MVOL)	229.5+/-7.0 cm**3/mol	20 deg C 760 Torr	(1)
Molecular Weight (MW)	350.35		(1)
PKA (PKA)	0.71+/-0.40	Most Basic 25 deg C	(1)
Polar Surface Area (PSA)	121.32 A**2		(1)

(1) Calculated using Advanced Chemistry Development (ACD/Labs) Software V8.14
(C) 1994-2009 ACD/Labs

Results

- Beilstein: substance not known
- Literature full-text databases including ACS :
Structure search not possible
Alternative text search based on wording combinations “thiazolo*” and “melting point”, “mp” delivered - no result

Conclusion: No melting point data available

Full-text reference of CAS -search

We ordered the original reference from the CAS search and if you look into this original reference:

Novel One-Step Synthesis of Thiazolo[3,2-*b*]1,2,4-triazoles

Alan R. Katritzky, Alfredo Pastor, and Michael Voronkov

Peter J. Steel; *Org. Lett.*, 2000, 2 (4), pp 429–431

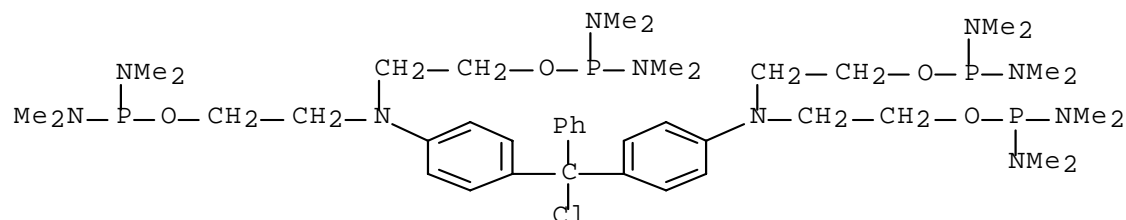
Table 1. Synthesis of Thiazolo[3,2-*b*]1,2,4-triazoles 5a–f

entry	R ¹	R ²	mp (°C)	yield (%)
5a	<i>p</i> -MeC ₆ H ₄	<i>p</i> -MeOC ₆ H ₄	140–141	22
5b	<i>p</i> -NO ₂ C ₆ H ₄	C ₆ H ₅	151–152	13
5c	<i>p</i> -MeC ₆ H ₄	<i>p</i> -ClC ₆ H ₄	162–163	20
5d	C ₆ H ₅	C ₆ H ₅	148–149	30
5e	<i>p</i> -MeC ₆ H ₄	C ₄ H ₃ S	oil	18
5f	<i>p</i> -ClC ₆ H ₄	C ₆ H ₅	154–155	12
6a	<i>p</i> -MeC ₆ H ₄	<i>p</i> -MeOC ₆ H ₄	134–135	31
6b	<i>p</i> -NO ₂ C ₆ H ₄	C ₆ H ₅	110–111	25
6c	<i>p</i> -MeC ₆ H ₄	<i>p</i> -ClC ₆ H ₄	82–83	19
6d	C ₆ H ₅	C ₆ H ₅	oil	12
6e	<i>p</i> -MeC ₆ H ₄	C ₄ H ₃ S	103–105	32
6f	<i>p</i> -ClC ₆ H ₄ CO	C ₆ H ₅	89–90	51

mp!!

Fulltext patent databases

Search for antihalation dye for FTO- search:



Structure search not possible in patent full text databases- work around by using wording

In 5 full text patent databases search in the fields title, abstract and full text : triarylmethane dye and phosph*

Or triarylmethane and phosph*

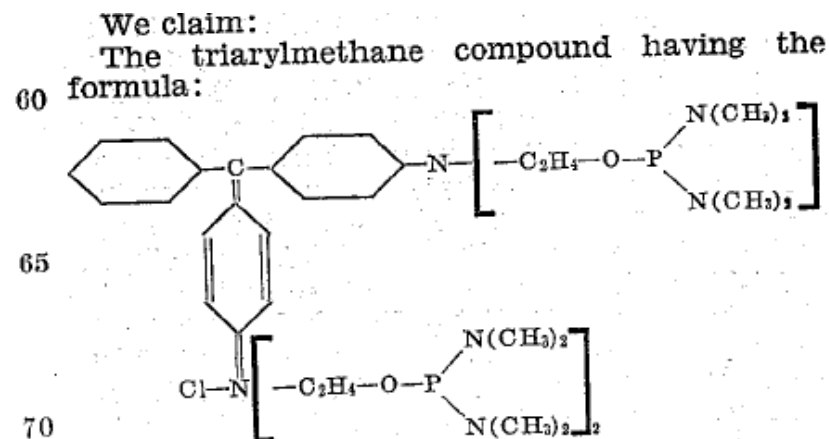
delivers several results but **not this antihalation dye**

☞ Structure based search in Chemical Abstracts delivers a result:
US2304946

Why did the search fail in patent full text databases:

1. Reference is fairly old – US2304946 was published 1942.
2. Most full-text databases only contained the title but no abstract and full-text for these old patents.
3. 1 database contained the full-text but OCR-error “triarylinethane” in place of triarylmethane
4. FTO-search based on claims would fail if the search was not done very broad

Claim 1 filed as:



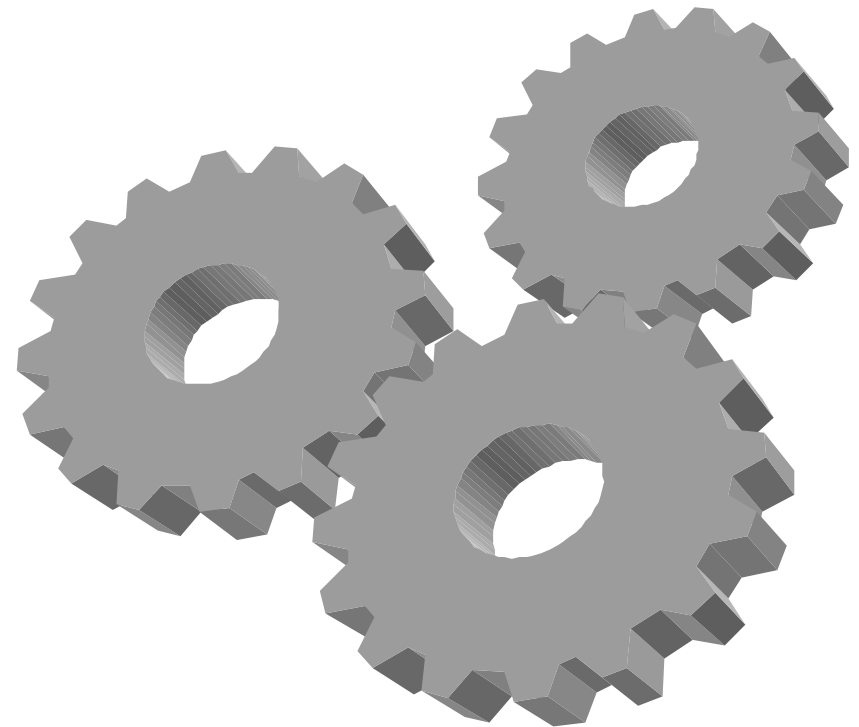
First Conclusions:

- **No result** in CAS, Beilstein or other indexed databases and in full-text databases for specific questions such as a specific property **does not mean that there is no result available**
 - ➔ Checking the original references is still needed
- **No result** in full-text database **does not** mean that there is no document
 - ➔ Indexed databases have added value in patent searching
 - ➔ Patent full-text databases may not cover all years and all countries

Lessons for the future

- Why do these searches fail?
- How can we avoid these pitfalls in our searches?

Let's have a look at the gear-wheels of a search

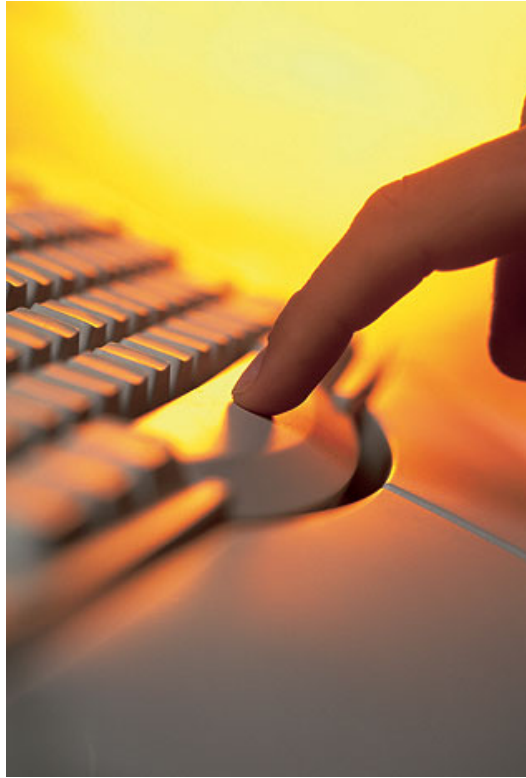


The gear-wheels in a search:

- The searcher
- The database
- The data as such



The database

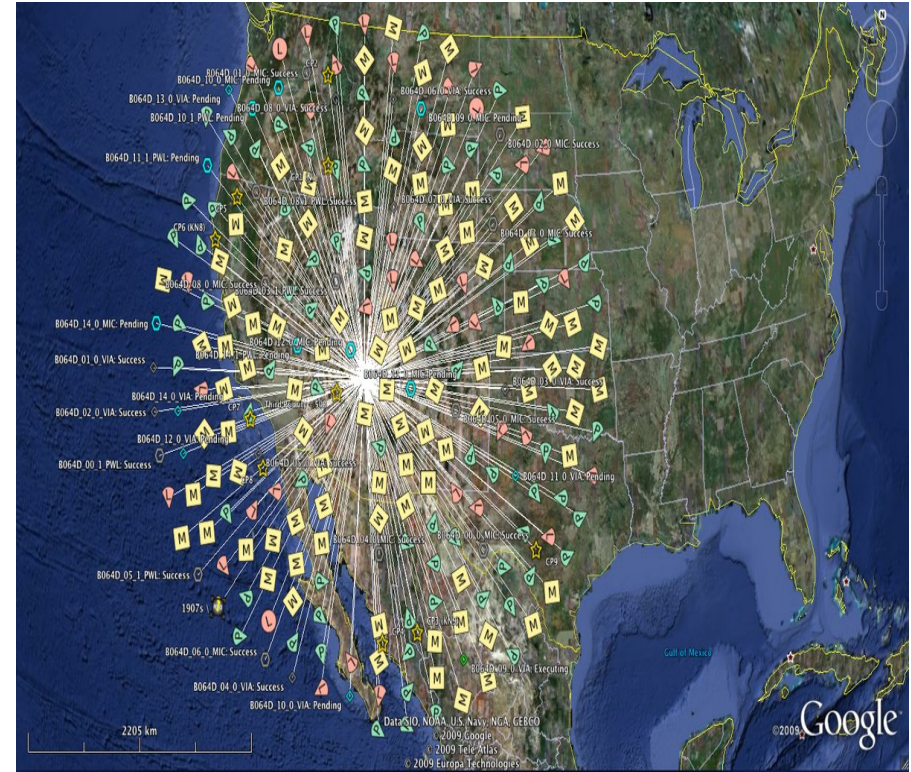


- **Scope of database**
(e.g. covered titles, patent authority)
- **Technical facilities**
(e.g. available search fields)
- **Quality of data**
(e.g. OCR errors, standardized bibliography)
- **Availability of data**
(e.g. patent authorities, electronic journals)

The data

2 Criteria :

- Type of data
 - Textual information
 - Non-textual information= Image Information
- Format of data
 - Electronic data
 - Electronically readable data
 - Non – electronic data



Textual Information: Pitfalls

Checklist before a search:

- Coverage of the provider (Journal-title, Patent-authority)
- Time coverage (poor availability of full - text before 1980; very poor access to data before 1970)
- Access to specific literature – is the provider covered by your company contracts?
- Which fields are searchable?
- Books (Are e-books available for your domain?)
- Conferences (Many conference content for application fields is not covered by any database)

Our failed search on parameters

To what type of data did the parameter in the table belong?

Obviously a to “Non-textual data: Image information”
and

Not a electronically readable digital image!

Table 1. Synthesis of Thiazolo[3,2-*b*]-1,2,4-triazoles 5a–f

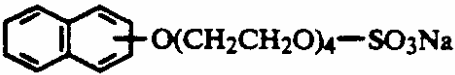
entry	R ¹	R ²	mp (°C)	yield (%)
5a	<i>p</i> -MeC ₆ H ₄	<i>p</i> -MeOC ₆ H ₄	140–141	22
5b	<i>p</i> -NO ₂ C ₆ H ₄	C ₆ H ₅	151–152	13
5c	<i>p</i> -MeC ₆ H ₄	<i>p</i> -ClC ₆ H ₄	162–163	20
5d	C ₆ H ₅	C ₆ H ₅	148–149	30
5e	<i>p</i> -MeC ₆ H ₄	C ₄ H ₃ S	oil	18
5f	<i>p</i> -ClC ₆ H ₄	C ₆ H ₅	154–155	12
6a	<i>p</i> -MeC ₆ H ₄	<i>p</i> -MeOC ₆ H ₄	134–135	31
6b	<i>p</i> -NO ₂ C ₆ H ₄	C ₆ H ₅	110–111	25
6c	<i>p</i> -MeC ₆ H ₄	<i>p</i> -ClC ₆ H ₄	82–83	19
6d	C ₆ H ₅	C ₆ H ₅	oil	12
6e	<i>p</i> -MeC ₆ H ₄	C ₄ H ₃ S	103–105	32
6f	<i>p</i> -ClC ₆ H ₄ CO	C ₆ H ₅	89–90	51

These different types of image information may result in a failing search:

- Tables
- Diagrams
- Figures and Drawings
- Flow-charts
- Photographs, Pictures
- others

Image data :Tables

Amino acid linked At 4-OH group of Example 1 compound	Salt	Yield %	Melting Point °C	MASS (M+H)	HPLC Purity (%)	Moisture Content
25. S-Ala	HCl, H ₂ O	95	160-5(d)	432	99.0	3.4
26. R-Ala	HCl, H ₂ O	92	225	432	99.0	4.7
27. S-Ala-S-Ala	HCl, 0.5 H ₂ O	90	190-93	503	97.0	1.1
28. R-Leu	HCl, H ₂ O	94	220-33	474	99.0	3.4
29. N-Me-S-Ala	HCl, 0.5 H ₂ O	91	140-50	446	97.5	1.9
30. R-Ala	AcOH	98	125-27	432	99.7	-
31. S-Val	HCl, 0.75 H ₂ O	93	160-61	460	96.2	2.7
32. S-Ala-S-Ala	HCl	60	175-80	503	97.8	-
33. R-Ala-R-Ala	HCl	75	95-100	503	98.0	-
34. S-Arg (Nitro)	HCl	70	113-16	588	93.0	-
35. S-Arg	HCl	70	178-82	603	94.0	-
36. [S-Arg(Nitro)] ₂	HCl					
37. S-Arg-S-Arg	3HCl					

Developer No.	Active Substance(s)	Amount of wt. %
1 (C)	Potassium silicate	7.6
2 (C)	Potassium silicate	8.6
3* (C)	Potassium silicate plus	8.6
		
4** (C)	Potassium silicate plus	8.6
	C ₁₂ H ₂₅ -O-CH ₂ -CH ₂ -OCH ₂ COOH	0.2
5*** (C)	Potassium silicate plus	8.6
	Triton ® H66 (Röhm & Haas) plus	2.5
	Synperonic ® T 304 (ICI) (ethylene oxide/propylene-oxide copolymer)	0.2
6*** (C)	as 5(C), but without Triton ® H66	
7	Potassium silicate plus O,O'-bis-carboxymethyl-polyglycol-600	8.6 0.5
8	Potassium silicate plus O,O'-bis-carboxymethyl-polyglycol-1000	8.6 0.5

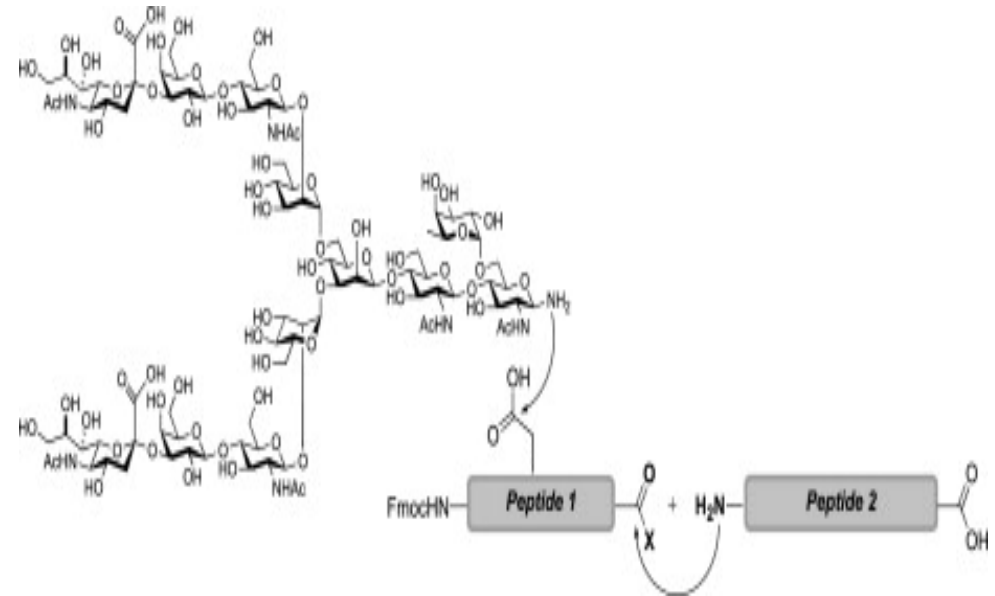
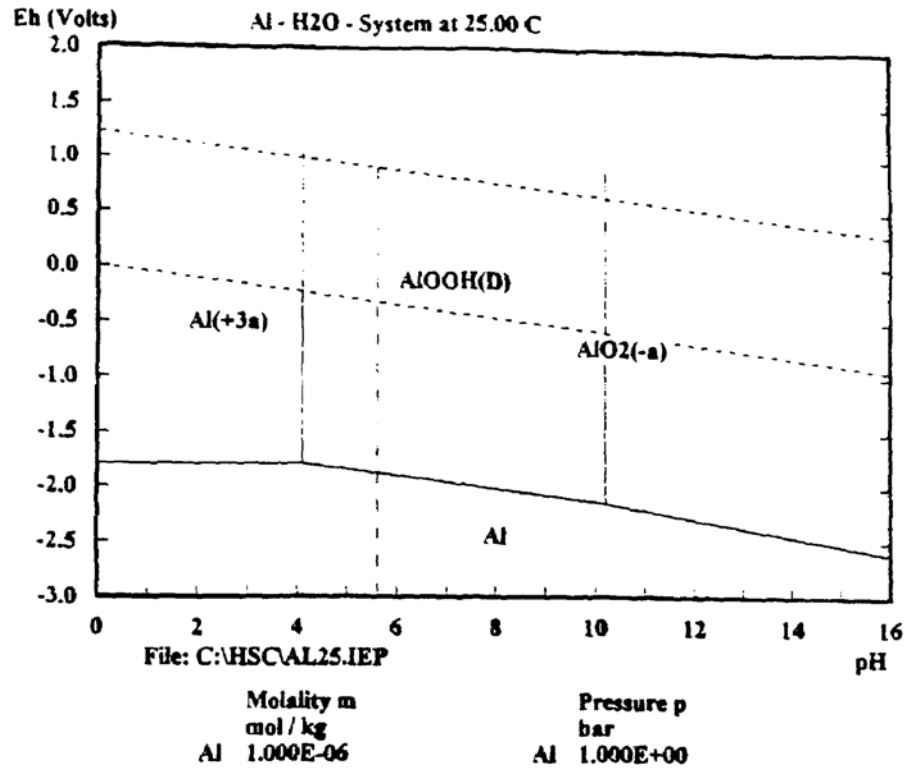
(C) Comparison

*according to DE 39 15 141

**according to EP 178 496

***according to EP 274 044

Image data: Diagrams



Graphical Abstract from Tetrahedron

Image data : Figures and Drawings

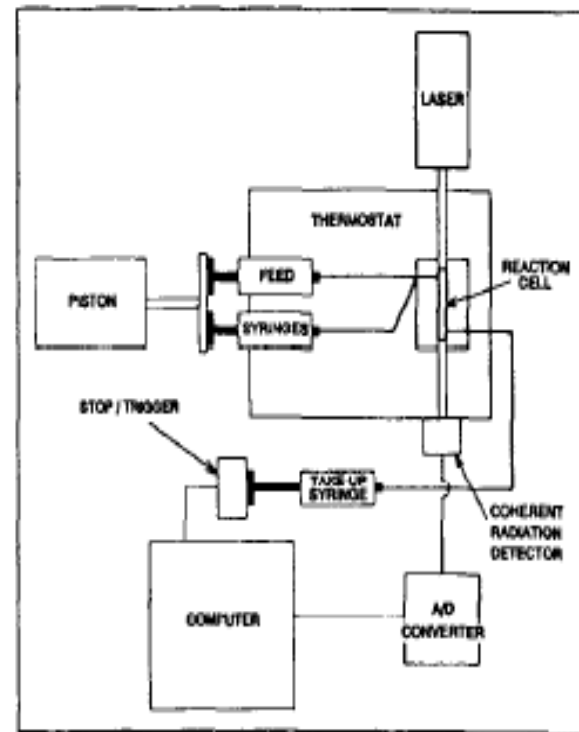
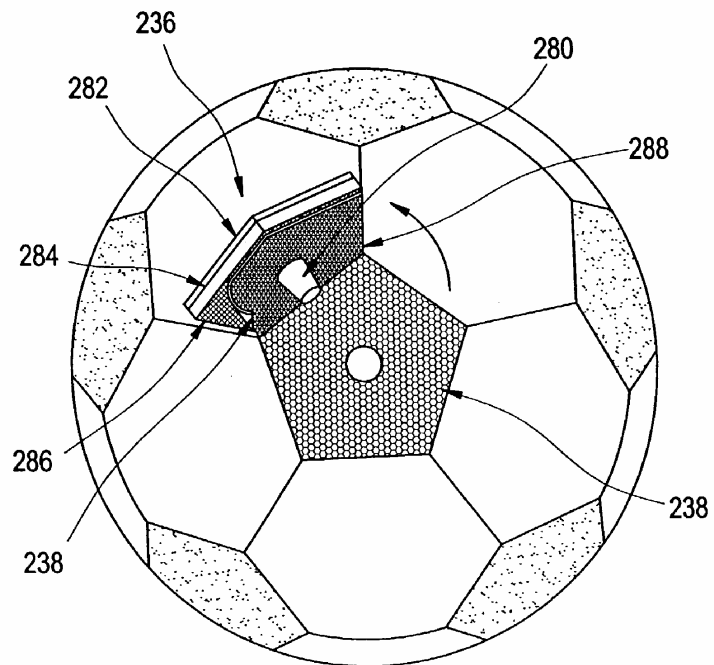


Fig. 1. Schematic diagram of the stopped flow apparatus.

Image data: Flow-charts

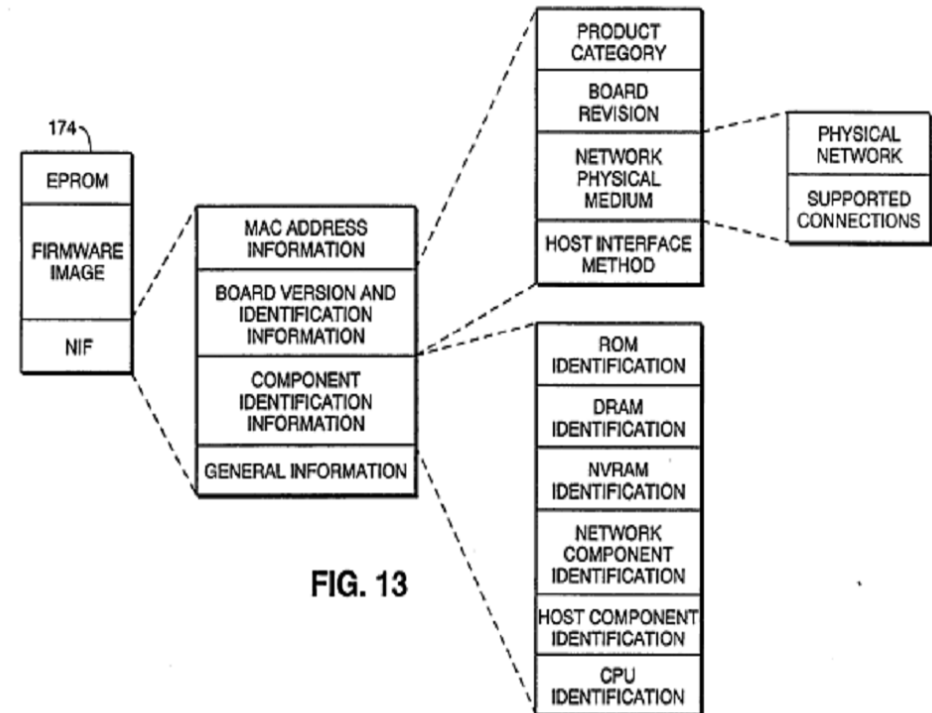
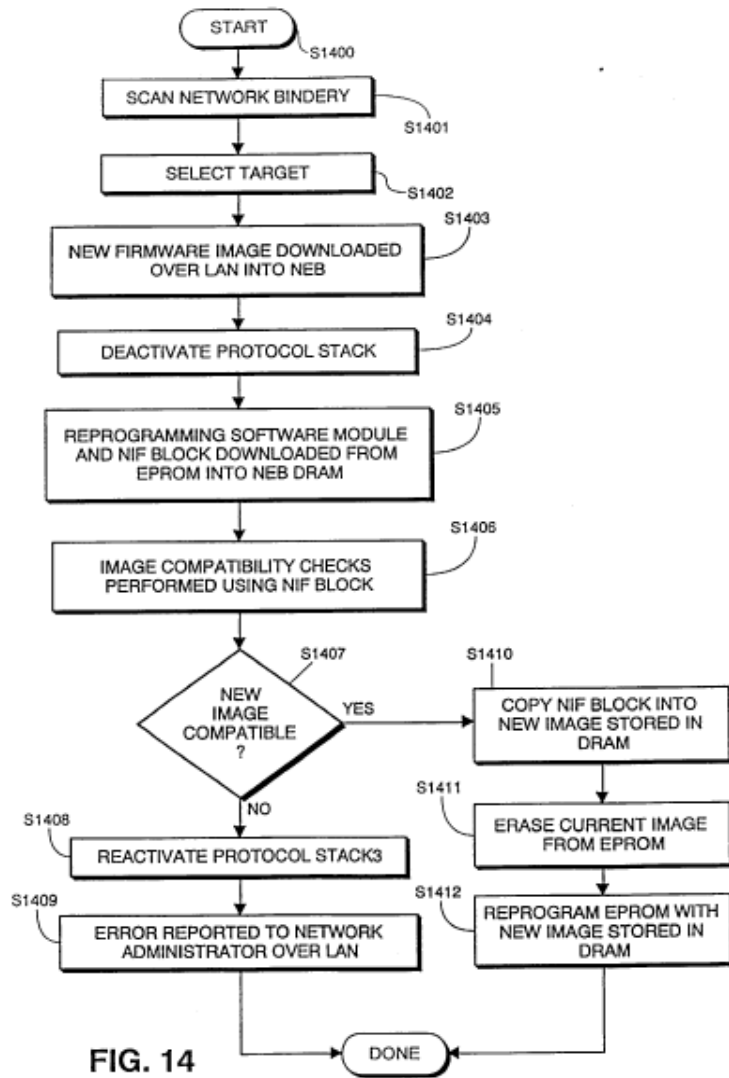


Image data : other picture information

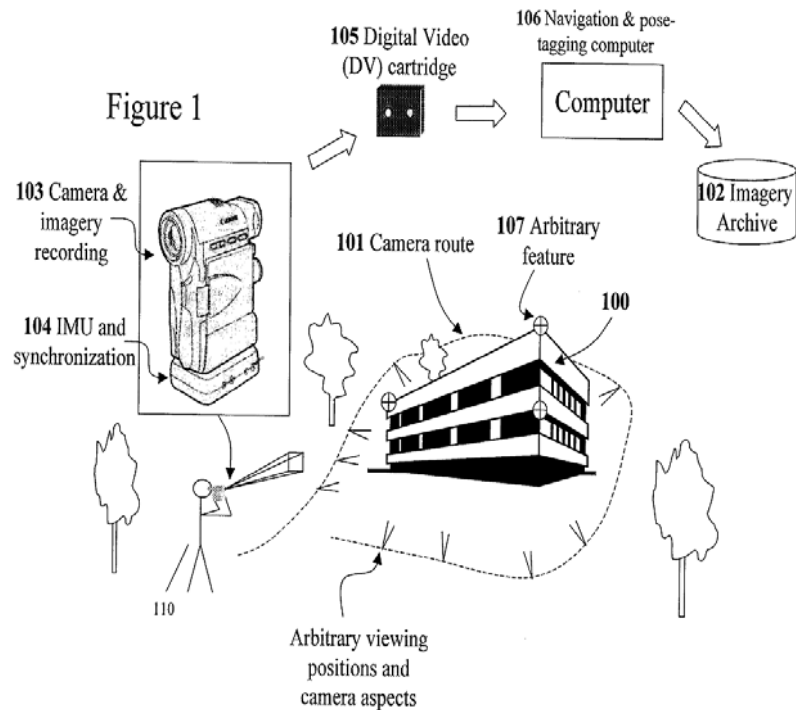


Fig. 1. Stages of development at which flowers were harvested. Left, Stage A; centre, Stage B; right, Stage C.

See: Cold storage of rose flowers: Effects of cold storage and water loss on opening and vase life of 'Mercedes' roses

J.D. Faragher₁, _a, S. Mayaka_a, T. Tirosha_a and A.H. Halevy_a

Type of data - Retrieval of Data

Textual information

Search options : 

Fulltext databases

Abstracting databases

Indexed databases

Non-textual data: Image information

Search options: 

???

Work-around with

Indexed databases

Can we expect improvements?

- Tables:
 - Simple tables are searchable in some ACS full-texts
 - Information retrieval research groups are reporting positive results

But problems expected with

- ☞ Amount of data to re-index
- ☞ Nested tables or image (Formula) in tables

- Images: S. Vrochidis reported successful results

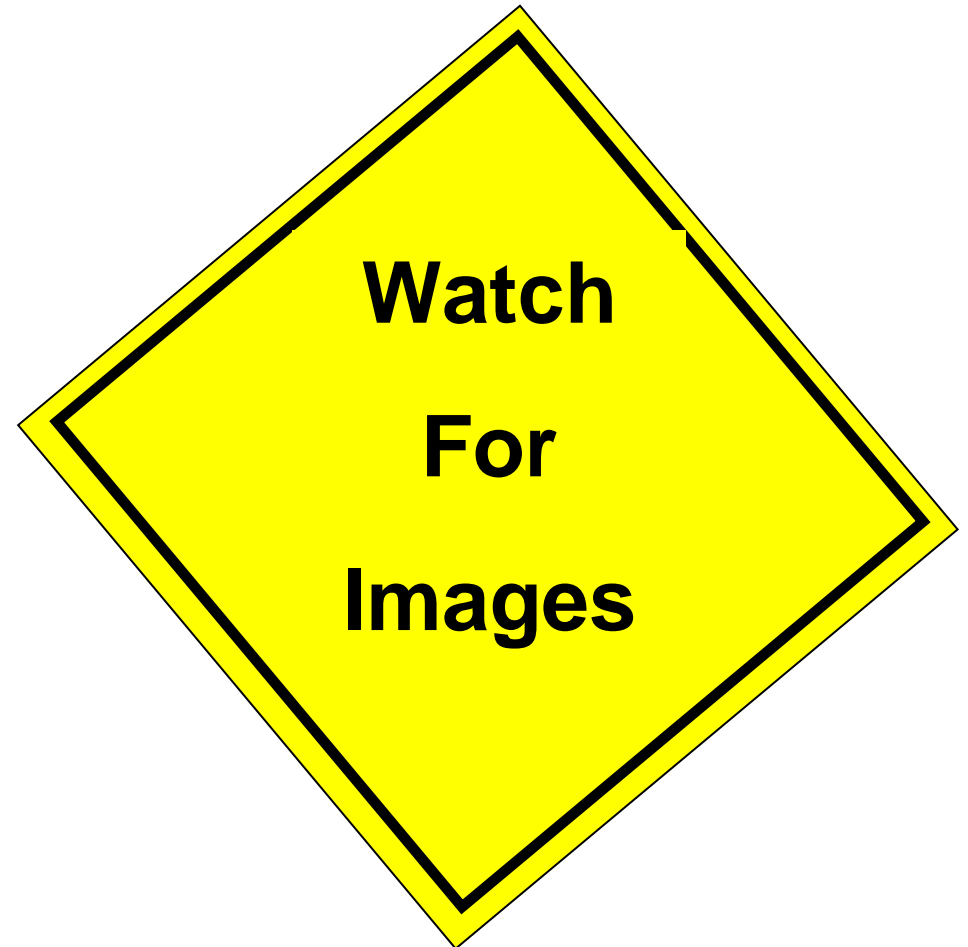
But problems expected with

- ☞ Amount of data to re-index
- ☞ Amount different types of image data

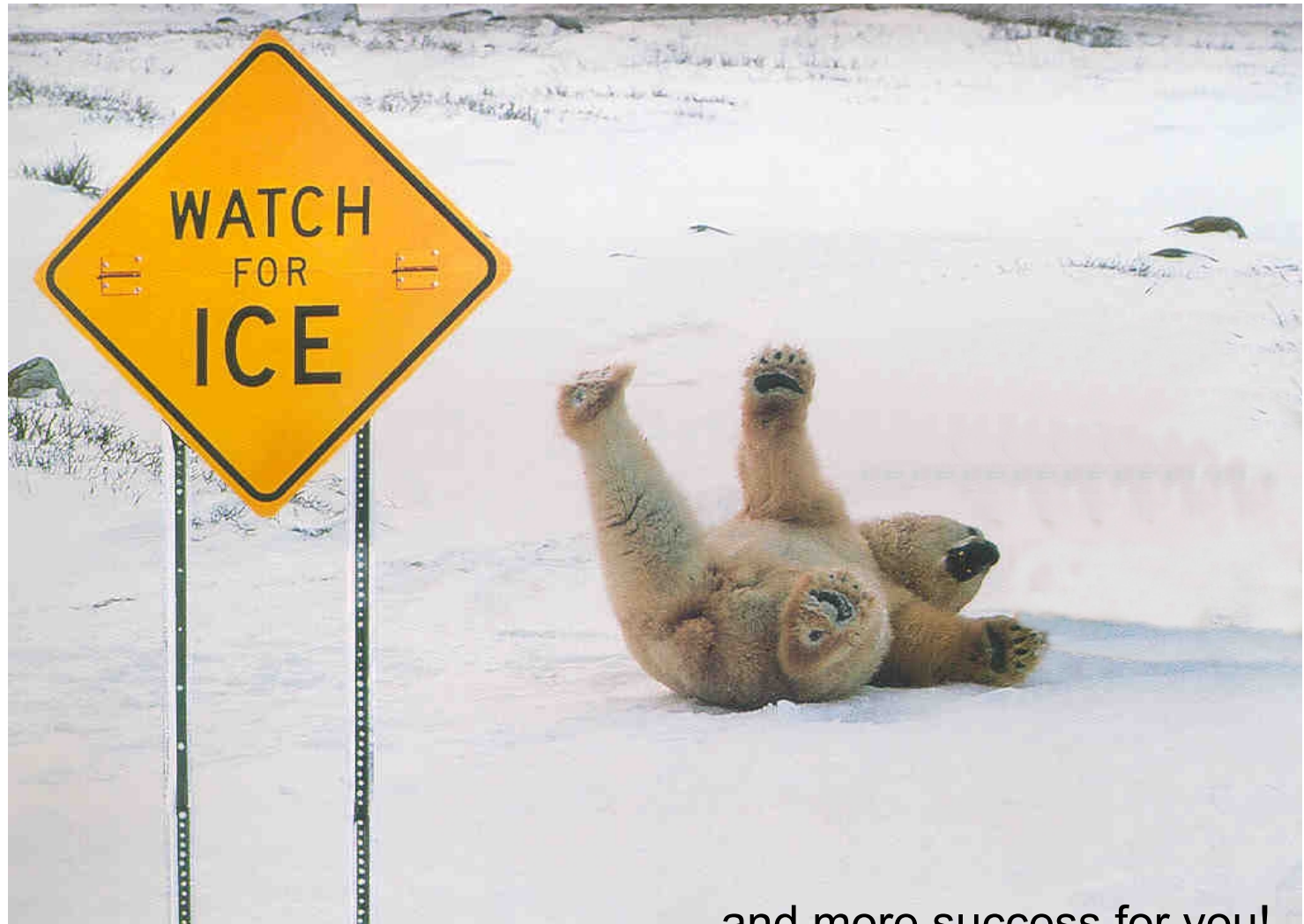
What to remember

Avoid slips by searching information hidden in these data formats:

- ✓ Tables
- ✓ Diagrams
- ✓ Figures and Drawings
- ✓ Flow-charts
- ✓ Photographs, Pictures
- ✓ others



Pay attention



- and more success for you!