

# *Cross-Language Information Retrieval in the Legal Domain*

*22.7.2010, Thomas Arni & Samir Abdou*



Overview:

- The Federal Supreme Court and its Decisions
- In Many Aspects a Special Retrieval Application
  - Specialty of Documents
  - DB-Queries vs. IR-Queries
  - Targeting Specific User Groups
- The Role of Cross-Language Information Retrieval (CLIR)
  - CLIR is vital
  - Thésaurus de Droit Suisse (TDS)
  - Quantity vs. Quality
- The Role of User Guidance
  - Revealing the existence of decisions (more recent, other version, other language)
  - Suggesting norms and descriptors
- Summary and Lessons Learned

- Since 1875 in Lausanne
- Supreme Jurisdiction in Switzerland  
(not a constitutional court)
- Publishes Decisions which
  - address a new issue or
  - overrule an old decisions
- The language of the decision is either German, French, Italian, or Romansh (Rumantsch Grischun)
- The decisions (approx. 90'000) 1954 – 2010 are online and searchable since 1997



Copyright [http://de.wikipedia.org/wiki/Benutzer:Roland\\_Zumb%C3%BChl](http://de.wikipedia.org/wiki/Benutzer:Roland_Zumb%C3%BChl)



# Structure of a Decision

## Urteilkopf

136 II 5

2. Auszug aus dem Urteil der II. öffentlich-rechtlichen Abteilung i.S. X. und Y. gegen Sicherheitsdirektion und Regierungsrat des Kantons Zürich (Beschwerde in öffentlich-rechtlichen Angelegenheiten)  
2C\_496/2009 vom 29. September 2009

## Regeste

**Art. 7 lit. d FZA** sowie **Art. 3 und 5 Anhang I FZA**; Nachzug von Familienmitgliedern mit Drittstaatsangehörigkeit; Anpassung der bundesgerichtlichen Rechtsprechung an die geänderte Rechtsprechung des Gerichtshofes der Europäischen Gemeinschaften; Voraussetzungen, unter denen eine Bewilligung verweigert werden kann.

- Prozessuales (E. 1).
- Tatsächliches (E. 2).

Regelung des Familiennachzuges nach dem Freizügigkeitsabkommen (E. 3.1-3.3). Berücksichtigung der nachträglichen Entwicklung der Rechtsprechung des Gerichtshofes der Europäischen Gemeinschaften: Zur Gewährleistung einer parallelen Rechtslage zwischen den Mitgliedstaaten der Europäischen Gemeinschaft und zwischen derselben und der Schweiz kann nicht an der Voraussetzung festgehalten werden, dass sich das Familienmitglied mit Drittstaatsangehörigkeit, das ein Angehöriger eines EU-Staates in die Schweiz nachziehen will, vorgängig bereits rechtmässig in der Schweiz oder einem anderen Vertragsstaat aufgehalten hat (Änderung der Rechtsprechung gemäss **BGE 130 II 1** und **BGE 134 II 10** in Anpassung an das Urteil des EuGH i.S. Metock; E. 3.4-3.7).

Voraussetzungen des Freizügigkeitsabkommens für die Verweigerung einer Bewilligung, auf die Anspruch besteht, insbesondere Erfordernis einer gegenwärtigen und hinreichend schweren Gefährdung der öffentlichen Ordnung (Fortführung von **BGE 129 II 215** und **BGE 130 II 176**; E. 4).

Rechtsfolgen (E. 5.1)

## Sachverhalt ab Seite 7

BGE 136 II 5 S. 7

A.

**A.a** Der aus Palästina (Westbank) stammende X., geb. 1969, reiste am 25. Juni 1996 in die Schweiz ein, wo er um Asyl ersuchte. Das Gesuch wurde am 11. November 1998 abgewiesen. Am 8. September 1998 wurde X. wegen Betäubungsmitteldelikten mit 18 Monaten Gefängnis bei bedingtem Vollzug und mit einer Landesverweisung für die Dauer von sieben Jahren bestraft. Die in der Folge angeordnete Ausschaffungshaft führte nicht zum Ziel, da die Ausschaffung mangels Reisepapieren nicht vollzogen werden konnte. Der danach ergangenen Aufforderung, die Schweiz selbständig zu verlassen, kam X. nicht nach. Am 25. Mai 2000 verurteilte ihn das Obergericht des Kantons Zürich erneut wegen Betäubungsmitteldelikten sowie wegen falschen Zeugnisses zu zehn Monaten Gefängnis unbedingt. Gleichzeitig ordnete es den Vollzug der bedingt ausgefallenen Gefängnisstrafe vom 8. September 1998 an. Vom 19. September 1999 bis zum 12. November 2001 befand sich X. zuerst in Untersuchungshaft und später im Strafvollzug.

## Inhalt

Ganzer BGE  
Regeste: [deutsch](#) [französisch](#) [italienisch](#)

## Sachverhalt

Erwägungen [1](#) [2](#) [3](#) [4](#) [5](#)

[Druckansicht \(PDF\)](#)

## Publikationen

AJP 2010 102

## Kommentare

M. Spescha, AJP, 2010 104 (d)

## Intellektuelle Indexierung

### Normen

Aus Teilregeste a:

[173.110 \(2005\)](#), [83 \(2005\) c \(2005\) 2 \(2005\)](#)  
[142.20 \(2005\)](#), [126 \(2005\)](#)  
[0.101 \(1950\)](#), [8 \(1950\) 1 \(1950\)](#)  
[101 \(1999\)](#), [13 \(1999\) 1 \(1999\)](#)  
[0.142.112.681 \(1999\)](#), [7 \(1999\) d \(1999\)](#)  
[0.142.112.681 \(1999\)](#), [16 \(1999\) 2 \(1999\)](#)  
[0.142.112.681.S.1 \(1999\)](#), [3 \(1999\) 1 \(1999\)](#)  
[0.142.112.681.S.1 \(1999\)](#), [3 \(1999\) 2 \(1999\)](#)  
[1 \(1999\)](#)  
[0.142.112.681.S.1 \(1999\)](#), [5 \(1999\) 1 \(1999\)](#)  
[0.142.112.681.S.1 \(1999\)](#), [5 \(1999\) 2 \(1999\)](#)

### Deskriptoren

Aus Teilregeste a:

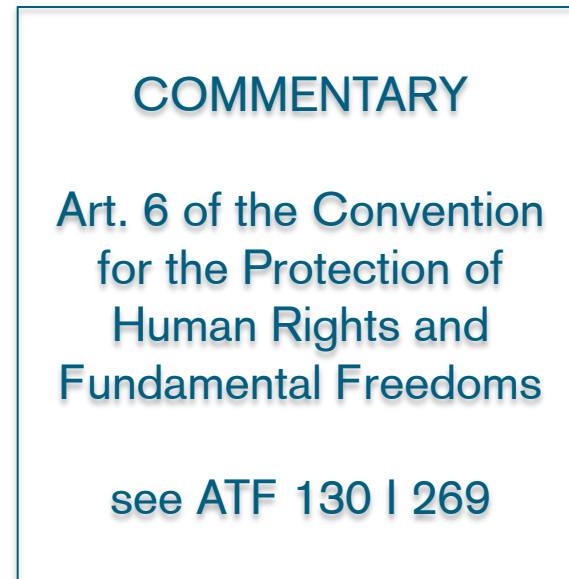
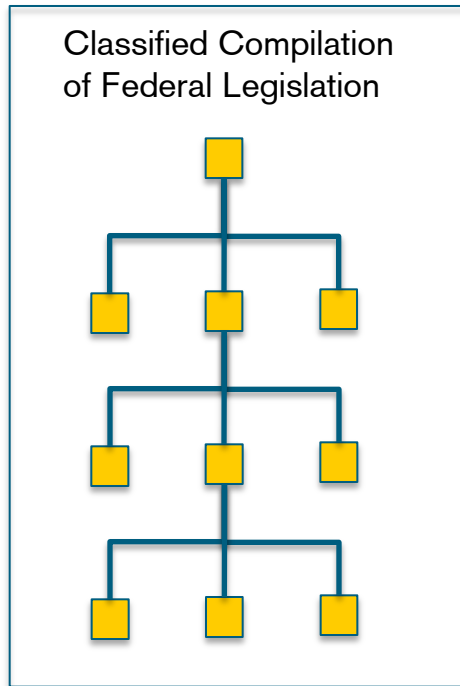
[RECHTSMITTEL ?](#)  
[BUNDESGERICHT ?](#)  
[BESCHWERDE IN ÖFFENTLICH-RECHTLICHEN ANGELEGENHEITEN ?](#)  
[ANFECHTUNGSGEGENSTAND ?](#)



Legal Domain IR	News Retrieval à la TREC/CLEF
Homogeneous in respect of quality of content and language	Heterogeneous
Usually non-overlapping cases	Many news cover the same content
Specific legal terms (decompounding often destroys meaning)	Everyday language
Structural and normalized data complements full text	Full text with almost no structure
Language may change within section	One language per document

## Database Queries vs. Information Retrieval Queries

As opposed to other domains, there is a well known hierarchical structure of the domain. Commentaries are widely used and refer directly to relevant decisions.



Between 20% to 30% of the queries are as easy a look up (DB queries).

- Language independence and the specificity of the references as well as different versions of enactments can be tricky.

The facts of a case are manually indexed by descriptors, but ....

- users may not know the descriptors of the thesaurus.
- the concept may be missing in the thesaurus, e.g.

- angemessenen Verfahrensdauer
- angemessene Dauer des Strafverfahrens
- Unangemessenheit
- Angemessenheit der Verfahrensdauer

Hence ....

- high-end text retrieval (stemming, decomposing, weighting/ranking) is required, which is coping with the uncertainties very typical for **Information Retrieval queries**.

## *Well Defined User Groups*

---

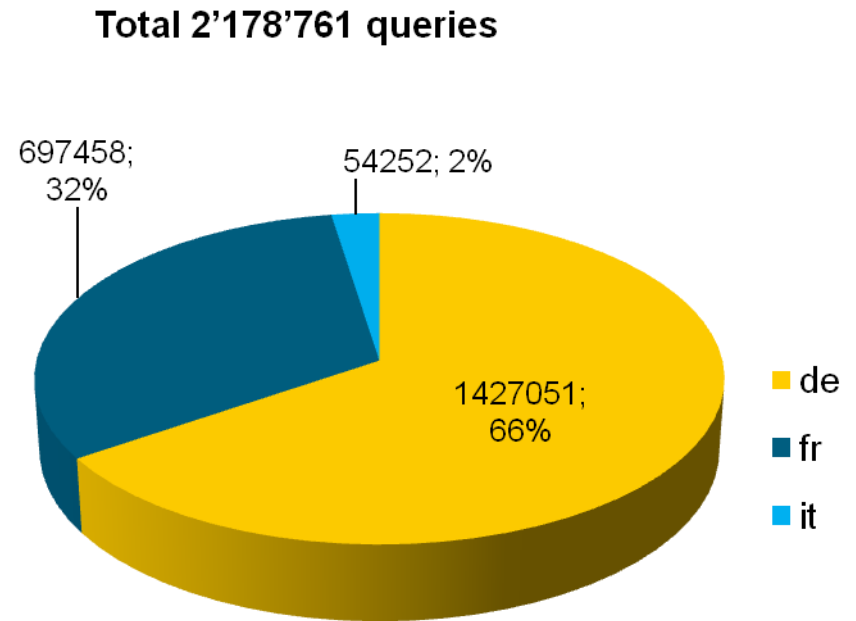
As opposed to many other retrieval applications,  
the user groups can easily be identified and characterized:

User Group	Recall	Precision	Versions
Specialists working on a specific case (judge at a court, state attorney, lawyer in a law firm)	++	0	+
Students in Law	+	++	0
Specialists in the Government publishing and producing new enactments	+	+	++



CLIR is crucial for retrieving decisions by the Federal Supreme Court:

- In contrast to the Swiss law, the decisions by the Federal Supreme Court are not translated.
- User queries are multilingual
  - German 66%
  - French 32%
  - Italian 2%
- Lawyers must know the decisions relevant to their case, it is a major duty to take care.



To accomplish well working CLIR is

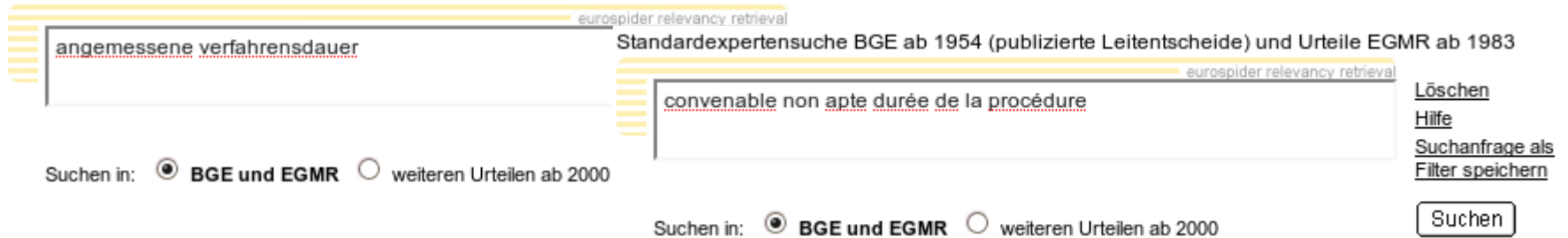
- is rather easy for less than 50% of the queries (i.e. for the DB queries).
- is challenging for more than 50% of the queries!

# The Role of Cross-Language Information Retrieval (CLIR)

CLIR is based on a multistep approach:

- monolingual search (source language is determined heuristically).
- pseudo-translation of the original query (user may modify the translation)
- monolingual search in the target language

Standardexpertensuche BGE ab 1954 (publizierte Leitentscheide) und Urteile EGMR ab 1983



The screenshot shows two search input fields. The top field contains the German query "angemessene verfahrensdauer" and is labeled "Standardexpertensuche BGE ab 1954 (publizierte Leitentscheide) und Urteile EGMR ab 1983". The bottom field contains the pseudo-translated French query "convenable non apte durée de la procédure" and is labeled "Standardexpertensuche BGE ab 1954 (publizierte Leitentscheide) und Urteile EGMR ab 1983". Both fields have a "eurospider relevancy retrieval" label above them. Below the top field, there are radio buttons for "Suchen in: BGE und EGMR" (selected) and "weiteren Urteilen ab 2000". To the right of the bottom field, there are links for "Löschen", "Hilfe", "Suchanfrage als Filter speichern", and a "Suchen" button.

Pseudo-translation by means of the Thésaurus de Droit Suisse (TDS), where

- TDS includes Jurivoc thesaurus used for manual indexing.
- multiword phrases are taken into account.
- TDS is annually updated with new terms.
- TDS-3 (2009) contains >360'000 terms.

## Experiences with the Thésaurus de Droit Suisse (TDS):

- Coverage of the query language (not necessarily document language) is more important than linguistically perfect translations.
- However, technical terms must be translated correctly (Geldbusse vs. Geldstrafe).
- Long multiword terms are highly desirable, even though they do not correspond to linguistic standard units (cf. translation memories).
- Periodic updates of the TDS with new terms is crucial.
- Automated update procedure is a major success factor of this CLIR system.

### TDS Entry : Geldbusse

#### Basic properties

Language: de  
Value: geldbusse  
Class: Noun

#### Unary relations

legal-term

#### Translations

fr AMENDE  
it MULTA

#### Binary relations

PAF GELD  
PAL BUSSE  
SBSF geldbussen  
USE BUSSE





The very domain specific objectives of the User Guidance are ...

- revealing the existence of several versions of an enactment, e.g. Art. 144 ZGB → either the version of 1907 (competent jurisdiction) or the version of 1998 (hearing).

Standardexpertensuche BGE ab 1954 (publizierte Leitentscheide) und Urteile EGMR ab 1983

Search interface showing the query "art. 144 zgb" in a search box. To the right are links: "Löschen", "Hilfe", "Suchanfrage als Filter speichern", and a "Suchen" button. Below the search box, there are radio buttons for "Suchen in:  BGE und EGMR  weiteren Urteilen ab 2000".

30 exakte Treffer gefunden für **art. 144 zgb**

- 1. 133 III 553**  **PDF**  
71. Auszug aus dem Urteil der II. zivilrechtlichen Abteilung i.S. X. gegen Y. (Berufung) 5C.316/2006 vom 5. Juli 2007  
Regeste [\[D, E, I\]](#) Kindesanhörung (Art. 144 Abs. 2 ZGB). Werden Kinder durch eine beauftragte Drittperson angehört, muss diese unabhängig und qualifiziert sein. Wo dies für das Kind eine unzumutbare Belastung bedeuten würde, ist von einer erneuten Anhörung durch den Richt...
- 2. 131 III 553**  **PDF**  
71. Auszug aus dem Urteil der II. Zivilabteilung i.S. X. gegen Y. (Berufung) 5C.63/2005 vom 1. Juni 2005  
Regeste [\[D, E, I\]](#) Anhörung der Kinder (Art. 144 Abs. 2 ZGB). Im Sinn einer Richtlinie ist die Kinderanhörung grundsätzlich ab dem vollendeten sechsten Altersjahr möglich (E. 1).
- 3. 113 II 102**   
19. Auszug aus dem Urteil der II. Zivilabteilung vom 26. März 1987 i.S. X. gegen Y. (Berufung)  
Regeste [\[D, E, I\]](#) Ergänzung eines ausländischen Scheidungsurteiles (Zuweisung der elterlichen Gewalt); Gerichtsstand. Für die Ergänzung eines Scheidungsurteiles ist grundsätzlich der Scheidungsrichter zuständig; eine Ausnahme ist jedoch dann gegeben, wenn der ausländische...
- 4. 111 II 401**   
80. Auszug aus dem Urteil der II. Zivilabteilung vom 12. Dezember 1985 i.S. X. gegen Z. (Berufung)

**Normen (SR) (BGE ab 1990)**

Wollen Sie Normen mit Fassungen für Ihre Suche verwenden?

[<CH/210\[1907\]/144\[1907\]>](#)  
[<CH/210\[1907\]/144\[1998\]>](#)

**Deskriptoren (BGE ab 1990)**

Statistisch relevante Deskriptoren:

- [<EHESCHIEDUNG> ?](#)
- [<FAMILIENRECHT> ?](#)
- [<ANHÖRUNG DES KINDES> ?](#)
- [<ZUSTÄNDIGKEIT> ?](#)
- [<GÜTERTRENNUNG> ?](#)
- [<ZIVILPROZESS> ?](#)
- [<KIND> ?](#)
- [<ALTER> ?](#)

The very domain specific objectives of the User Guidance are ...

- suggesting norms to a full text query,  
e.g. „angemessene Verfahrensdauer“ → Art.6 EMRK
- suggesting descriptors representing a standard legal concept,  
e.g. „hat Geld gewaschen“ → <Geldwäscherei>.

Standardexpertensuche BGE ab 1954 (publizierte Leitentscheide) und Urteile EGMR ab 1983

europider relevancy retrieval

hat geld gewaschen

[Löschen](#)  
[Hilfe](#)  
[Suchanfrage als Filter speichern](#)

Suchen in:  **BGE und EGMR**  weiteren Urteilen ab 2000

3 exakte Treffer gefunden für **hat geld gewaschen**

1. **111 Ib 319** 

58. Auszug aus dem Urteil der Anklagekammer vom 11. Oktober 1985 i.S. K. gegen Bundesamt für Polizeiwesen  
Regeste [D, E, I] Art. 16 des Europäischen Auslieferungsübereinkommens (SR 0.353.1), Art. 28 und 50 Abs. 1 IRSG; vorläufige Auslieferungshaft. 1. Formelle Anforderungen, denen ein Ersuchen um vorläufige Verhaftung genügen muss. Frist, innert der allfällige Verbesserungen...

2. **119 IV 59** 

11. Auszug aus dem Urteil des Kassationshofes vom 20. Januar 1993 i.S. H. gegen Generalprokurator des Kantons Bern (Nichtigkeitsbeschwerde)  
Regeste [D, E, I] Art. 305bis Ziff. 1 StGB; Verstecken von Drogengeld. Den Grundtatbestand der Geldwäscherei erfüllt jede Tathandlung, die geeignet ist, die Einziehung der Verbrechensbeute zu vereiteln (E. 2). Bedeutung des Randtitels für die Auslegung (E. 2b/cc). Das Ve...

**Deskriptoren (BGE ab 1990)**

Suchterm "geld" mit Deskriptor <GELD> ? ersetzen?

Statistisch relevante Deskriptoren:

- <GELDWÄSCHEREI> ?
- <HANDEL MIT BETÄUBUNGSMITTELN> ?
- <STRAFBESTIMMUNGEN DES BETMG> ?
- <GELD> ?
- <MARGINALIE> ?
- <GEHILFENSCHAFT> ?
- <ANLAGE(INVESTITION)> ?
- <HAFTBEFEHL> ?

The very domain specific objectives of the User Guidance are ...

- Revealing the existence of decisions in other languages, e.g. Translate Query to French.
- Revealing the existence of a more recent decision overruling an old decision, e.g. by sorting decisions about a legal aspect in decreasing order of date.

- Importance of multilingualism.
  - Coverage of query terms is more important than linguistically perfect translations
- Combining DB- and IR-Queries is challenging. Users expect either DB -or IR-Queries, but hardly both.
- Applying approved IR methods in a specific domain.
- Such a special Retrieval Application requires a sophisticated combination of sound retrieval techniques.
- Recall matters as opposed to web search services.
- Exploiting domain specific characteristics, e.g.
  - domain specific entities, terms
- Automatically created dictionaries are useful.

## *Additional Information*

---

Federal Supreme Court : <http://www.bger.ch>

Eurospider : <http://www.eurospider.com>

Eurospider Podcasts: <http://podcasts.eurospider.com>

Thésuarus de Droit Suisse : <http://tds.eurospider.com>

Classified Compilation of Federal Legislation : <http://www.admin.ch/ch/e/rs/rs.html>