



Ad Hoc Track Overview: The TEL and Persian Tasks

Giorgio M. Di Nunzio

University of Padua
Italy

dinunzio@dei.unipd.it

Nicola Ferro

University of Padua
Italy

ferro@dei.unipd.it

Carol Peters

ISTI-CNR, Area di Ricerca Pisa
Italy

carol.peters@isti.cnr.it



Participation

24 participants
14 countries

Participant	Institution	Country
chemnitz	Chemnitz University of Technology	Germany
cheshire	U.C.Berkeley	United States
geneva	University of Geneva	Switzerland
imag	Inst. For Infocomm Research	France
inaoe	INAOE	Mexico
inesc	INESC ID	Portugal
isi	Indian Statistical Institute	India
ixa	Univ. Basque Country	Spain
jhu-apl	Johns Hopkins University Applied Physics Lab	United States
karlsruhe	University of Karlsruhe	Germany
know-center	Knowledge Relationship Discovery	Austria
opentext	Open Text Corporation	Canada
tehran-IRDB	IR-DB Research Group	Iran
tehran-NLP	NLP-Software Engineering Grad. Lab	Iran
tehran-NLPDB	NLP-DB Research Group	Iran
tehran-NLPDB2	NLP-DB Group	Iran
tehran-SEC	School of Electrical Computing-1	Iran
twente	Univ. of Twente	Netherlands
ucm	Universidad Complutense de Madrid	Spain
ufrgs	Univ. Fed. do Rio Grande do Sul	Brazil
uniba	Universita' di Bari	Italy
unine	U.Neuchatel-Informatics	Switzerland
xerox	Xerox Reseearch - Data Mining	France
xerox-sas	Xerox SAS	Italy

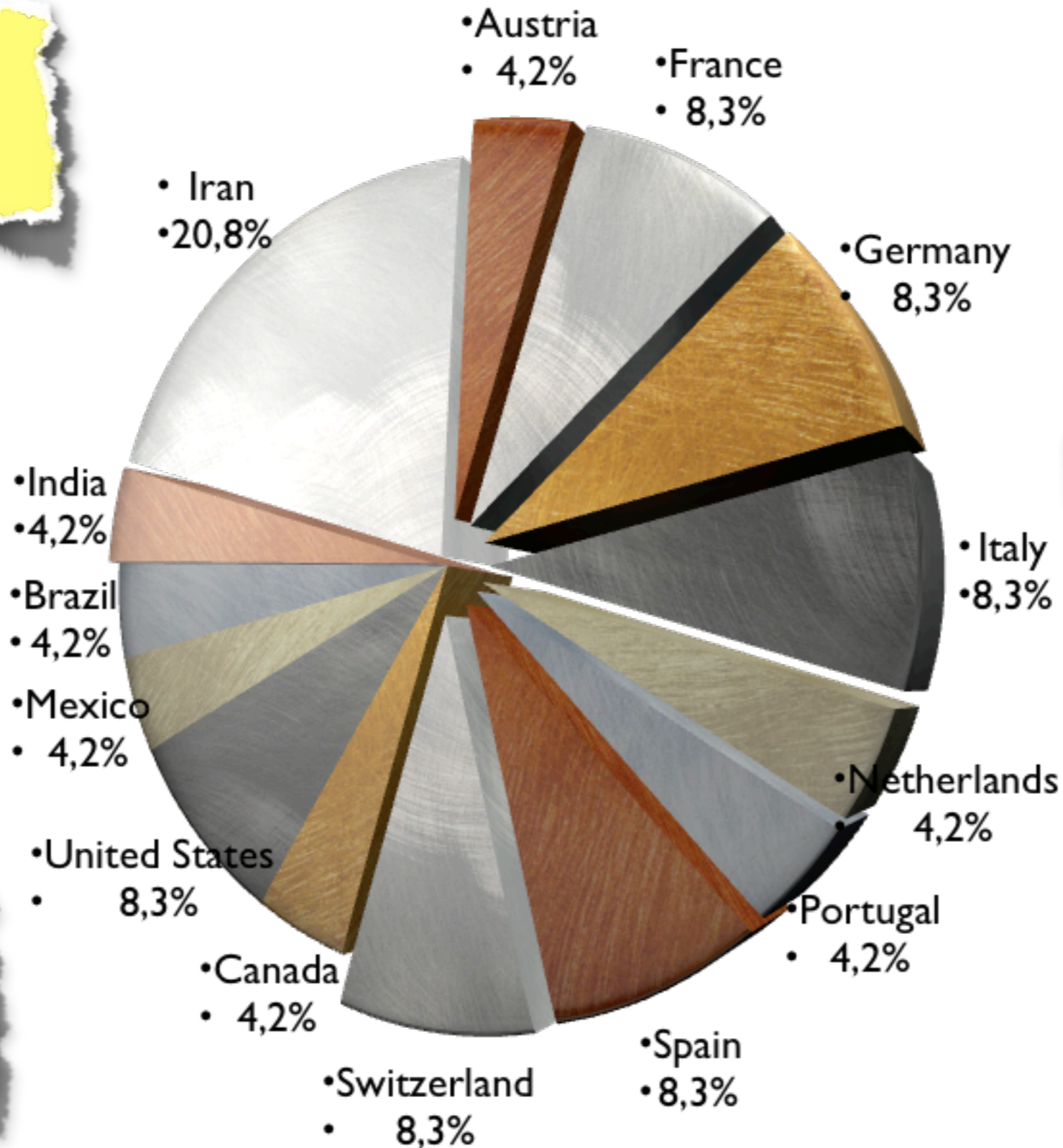


Participation by Country

**Asia:
25%**

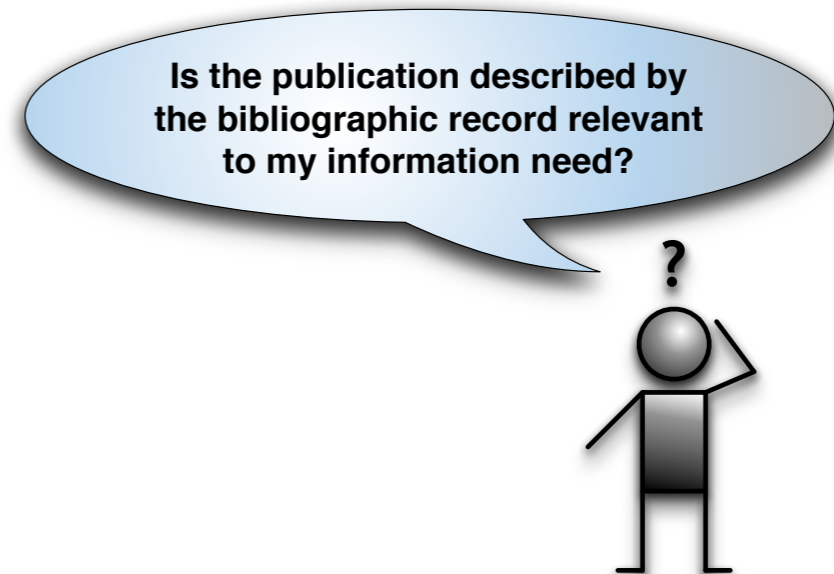
**Europe:
54%**

**America:
21%**



TEL Task

- ❑ The task is to search and retrieve relevant items from collections of **library catalog cards**, which are surrogates for documents held by libraries
- ❑ Both **monolingual** and **bilingual** tasks have been offered
- ❑ Not only the data are very sparse and less rich than newspapers but also the task is different from a traditional ad-hoc task





TEL Collections

- ❑ The collections have been provided by The European Library (<http://www.theeuropeanlibrary.org/>) and are catalog records harvested from Europe's national libraries

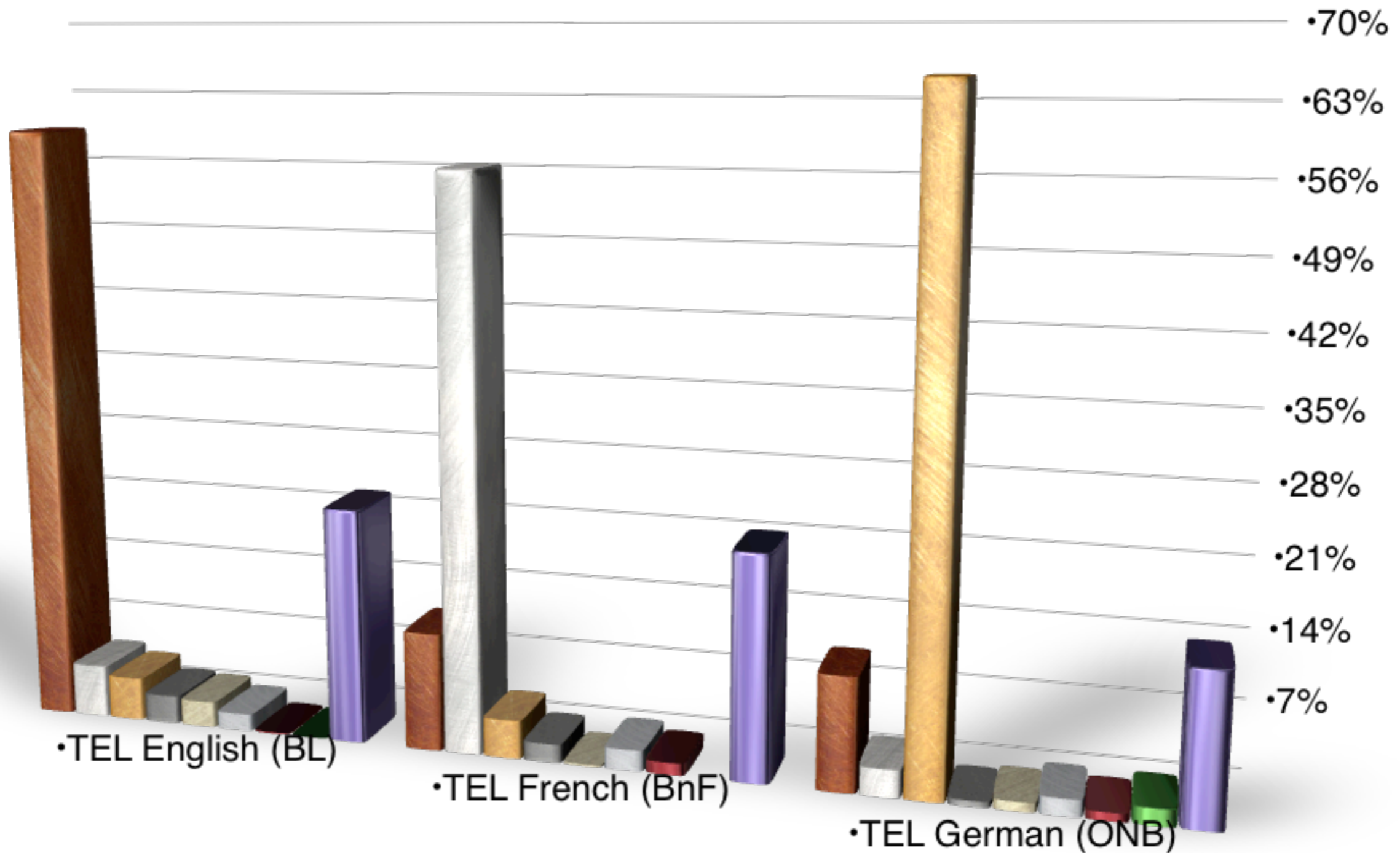
- ❑ English
 - source: British Library (BL)
 - size: 1,208,383,351 bytes
 - items: 1,000,100 records

- ❑ French
 - source: Bibliothèque Nationale de France (BnF)
 - size: 1.362.122.091 bytes
 - items: 1,000,100 records

- ❑ German
 - source: Austrian National Library (ONB)
 - size: 1.306.492.248 bytes
 - items: 869,353 records



TEL Collections: Distribution of the Languages

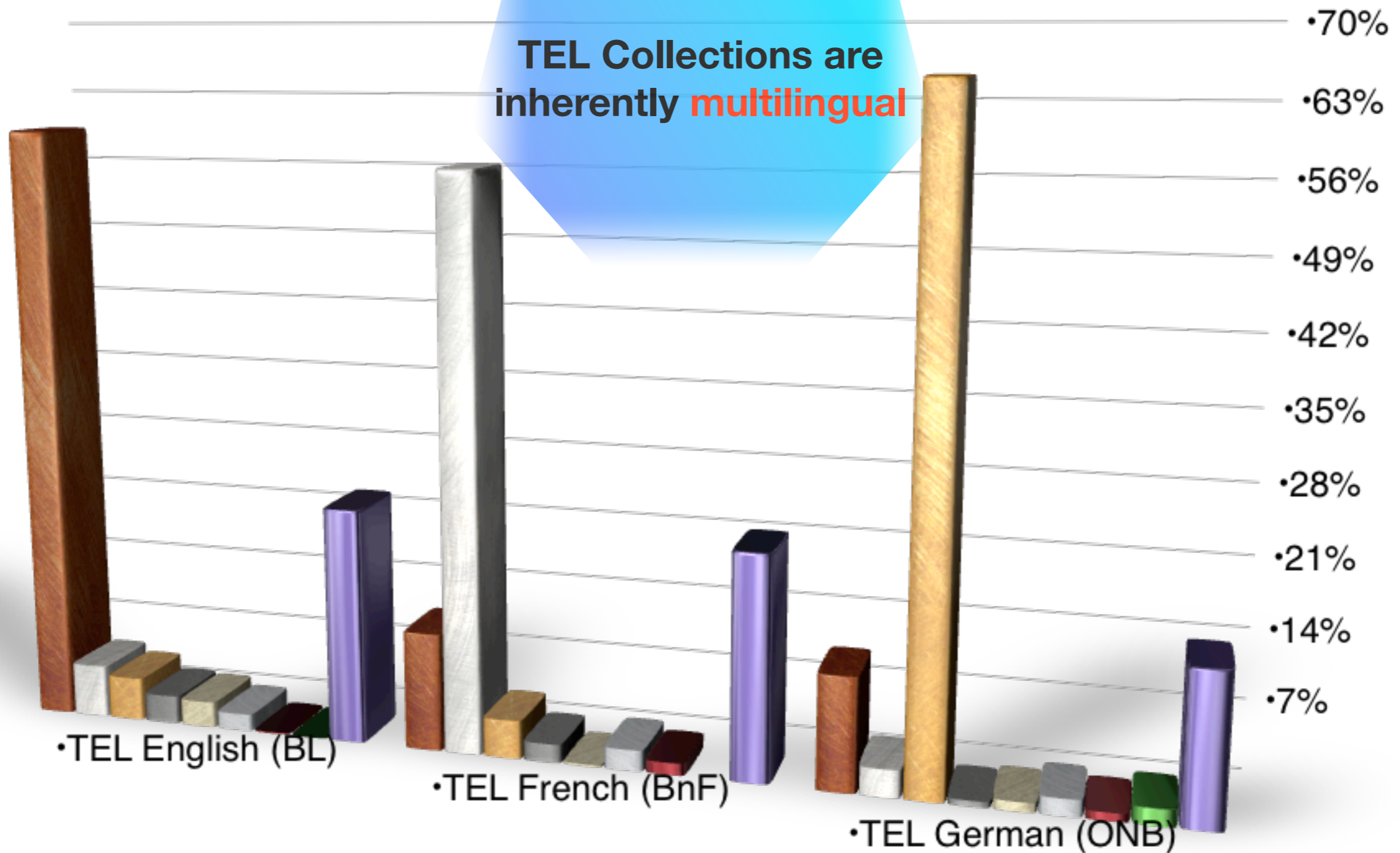




TEL Collections: Distribution of the Languages

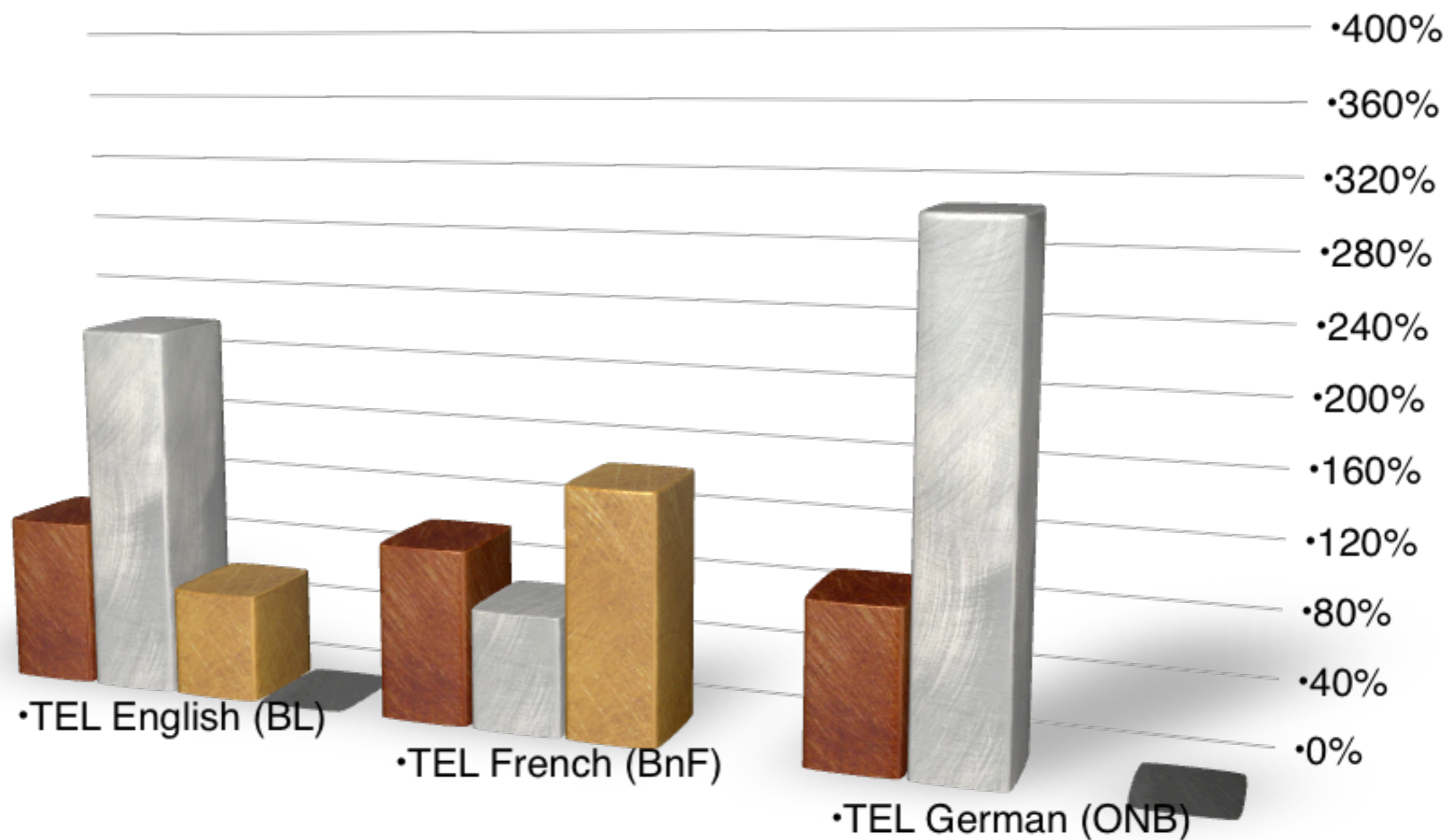


TEL Collections are inherently multilingual





TEL Collections: Distribution of the Content





TEL Collections: English BL Sample Record

```
<record>
  <set>TEL_BL_opac</set>
  <header>
    <id>010734316</id>
  </header>
  <document format="index">
    <index>
      <topic>BL_opac</topic>
    </index>
  </document>
  <document format="dcx">
    <oai_dc:dc>
      <dc:title>Country wives /</dc:title>
      <dc:contributor> Shaw, Rebecca, 1931- </dc:contributor>
      <dc:publisher>London : Orion, c2001</dc:publisher>
      <dc:terms:issued>c2001</dc:terms:issued>
      <dc:terms:extent>263 p. : 1 map ; 20 cm.</dc:terms:extent>
      <dc:language xsi:type="IS0639-2">eng</dc:language>
      <dc:description>Originally published: 2001.</dc:description>
      <dc:abstract>Country Wives is the second in a series of novels set in a veterinary practice on the outskirts of a Dorset town, by the bestselling author of the Turnham Malpas novels.</dc:abstract>
      <dc:subject xsi:type="dcterms:DDC">823.914</dc:subject>
      <dc:subject>Veterinarians</dc:subject>
      <dc:identifier xsi:type="lib:ISBN">0752844733 (pbk.) :</dc:identifier>
      <dc:type>text</dc:type>
      <dc:identifier
        xlink:href="http://catalogue.bl.uk/F/-?func=direct-doc-set& ;l_base=BLL01& ;from=TELgateway& ;doc_number=010734316"
        >010734316</dc:identifier>
      <dc:identifier xsi:type="dcterms:URI"
        >http://catalogue.bl.uk/F/-?func=direct-doc-set& ;l_base=BLL01& ;from=TELgateway& ;doc_number=010734316</dc:identifier>
      <mods:location>British Library HMNTS H.2002/3337</mods:location>
    </oai_dc:dc>
  </document>
</record>
```



TEL Collections: French BnF Sample Record

```
<record>
  <set>TEL_BnF_opac</set>
  <id>oai:bnf.fr:catalogue/ark:/12148/cb346737691/description</id>
  <document format="index">
    <index>
      <topic>BnF_opac</topic>
    </index>
  </document>
  <document format="dcx">
    <oai_dc:dc xmlns:oai_dc="http://www.openarchives.org/OAI/2.0/oai_dc/"
      xmlns:xsi="http://www.w3.org/2001/XMLSchema-instance"
      xmlns="http://www.openarchives.org/OAI/2.0/" xmlns:dc="http://purl.org/dc/elements/1.1/">
      <dc:identifiant>http://catalogue.bnf.fr/ark:/12148/cb346737691/description</dc:identifiant>
      <dc:title>Chine : la terre, l'eau et les hommes / écrit par Han Suyin ; photographié par
        Claude Sauvageot</dc:title>
      <dc:creator>Han, Suyin (1917-....)</dc:creator>
      <dc:contributor>Sauvageot, Claude (1935-....). Illustrateur</dc:contributor>
      <dc:publisher>Édition J.A. (Paris)</dc:publisher>
      <dc:date>1980</dc:date>
      <dc:description>Collection : Grands livres</dc:description>
      <dc:subject xml:lang="fre"> Chine -- Histoire</dc:subject>
      <dc:subject xml:lang="fre"> Chine -- Ouvrages illustrés</dc:subject>
      <dc:description>ISBN 2852581973</dc:description>
      <dc:language>fre</dc:language>
      <dc:type xml:lang="fre">texte imprimé</dc:type>
      <dc:type xml:lang="eng">printed text</dc:type>
      <dc:type xml:lang="eng">text</dc:type>
      <dc:rights xml:lang="fre">Catalogue en ligne de la Bibliothèque nationale de France</dc:rights>
      <dc:rights xml:lang="eng">French National Library online Catalog</dc:rights>
    </oai_dc:dc>
  </document>
</record>
```



TEL Collections: German ONB Sample Record

```
<record>
  <set>TEL_ONB_onb01</set>
  <id>oai:aleph.onb.ac.at:ONB01-000628794</id>
  <document format="index">
    <index>
      <topic>ONB_onb01</topic>
    </index>
  </document>
  <document format="dcx">
    <oai_dc:dc xmlns:oai_dc="http://www.openarchives.org/OAI/2.0/oai_dc/"
      xmlns:xsi="http://www.w3.org/2001/XMLSchema-instance"
      xmlns="http://www.openarchives.org/OAI/2.0/" xmlns:dc="http://purl.org/dc/elements/1.1/"
      xmlns:dcterms="http://purl.org/dc/dc/terms/"
      xsi:schemaLocation="http://www.openarchives.org/OAI/2.0/oai_dc/
http://www.openarchives.org/OAI/2.0/oai_dc.xsd">
      <dc:identifier xsi:type="onb:ACCRecordId">AC03970200</dc:identifier>
      <dc:language xsi:type="dcterms:ISO639-2">ger</dc:language>
      <dc:creator>Salema, Teresa</dc:creator>
      <dc:title>&lt;&lt;Des&gt;&gt; Widerspenstigen Zähmung in der
        Gesellschaft Wilhelm Meisters: Ordnung der Natur oder Ironie der Kultur</dc:title>
      <dcterms:issued>1986</dcterms:issued>
      <dc:identifier xsi:type="onb:CallNumber">834301-B.NEU-Per.C,31</dc:identifier>
      <dcterms:bibliographicCitation>&lt;&lt;Der&gt;&gt; Widerspenstigen
        Zähmung</dcterms:bibliographicCitation>
      <dc:publisher xsi:type="onb:PlaceofPublisher">Innsbruck</dc:publisher>
      <dcterms:issued>1986</dcterms:issued>
      <dcterms:bibliographicCitation>(1986), S.143 - 156</dcterms:bibliographicCitation>
      <dc:subject>Romantik</dc:subject>
      <dc:subject>Feministische Literaturwissenschaft</dc:subject>
      <dc:subject>Klassik</dc:subject>
      <dc:subject>Frauenbild</dc:subject>
      <dc:subject>Weiblichkeitsideal</dc:subject>
      <dc:subject>Rollenbild</dc:subject>
      <dc:subject>Goethe, Johann Wolfgang &lt;&lt;von&gt;&gt;; Wilhelm
        Meisters Lehr- und Wanderjahre</dc:subject>
      <dcterms:abstract>Die feministische Literaturwissenschaft der letzten Jahre hat sich
        erfolgreich bemüht, die patriarchalisch orientierten Denk- und Handlungsstrategien
        aufzudecken, die seit der Aufklärung der Befreiung des Subjekts als Citoyen und
        Bourgeois zugrundeliegen. Daß eine solche "Bfreiung" sowohl mit der Negierung der
        Frau als Subjekt als auch mit der Unterdrückung von für weiblichgehaltenen Werten im
        Mann zusammenhängt, ist bisher und nicht nur aus feministischer Perspektive
        weitgehend bestätigt worden. Daher erscheint es sinnvoll, nach der Notwendigkeit
        einer solchen "Zähmung" grundsätzlicher zu fragen, indem Alternativmodelle aus jener
        Zeit analysiert werden, die zwar einen Zivilisationsprozeß mit dazugehöriger
        Triebbändigung postulieren, dies jedoch nicht aus eindeutig sexistischer Sicht tun.
        Goethes "Wilhelm Meister" ist hier in doppelter Hinsicht ein positives
        Gegenbeispiel: seiner frauenfreundlichen Einstellung wegen, und auch weil die beiden
        Romane im Gegensatz zur Ltieratur der Romantik mit der Wirklichkeit in positiver
        Konfrontation stehen</dcterms:abstract>
    </oai_dc:dc>
  </document>
</record>
```



TEL Topics

```
<?xml version="1.0" encoding="UTF-8" standalone="no"?>
<topic>
  <identifier>10.2452/451-AH</identifier>

  <title lang="en">Roman Military in Britain</title>
  <title lang="de">Römisches Militär in Britannien</title>
  <title lang="es">El ejército romano en Britania</title>
  <title lang="fr">L'armée romaine en Grande-Bretagne</title>
  <title lang="nl">Romeinse Leger in Groot-Brittannie</title>

  <description lang="en">Find books or publications on the Roman invasion or military occupation
    of Britain.</description>
  <description lang="de">Finden Sie Bücher oder Publikationen über die römische Invasion oder das
    Militär in Britannien.</description>
  <description lang="es">Encuentre libros o publicaciones sobre la invasión romana o la ocupación
    militar romana en Britania.</description>
  <description lang="fr">Trouver des livres ou des publications sur l'invasion et l'occupation de
    la Grande-Bretagne par les Romains.</description>
  <description lang="nl">Vind boeken of publicaties over de Romeinse invasie of bezetting van
    Groot-Brittannie.</description>
</topic>
```

- ❑ 50 topics have been developed in English, German, and French
- ❑ Additional translations to Dutch and Spanish have been provided upon request
- ❑ Topics consist of **title** and **description** only; the narrative contained information relevant only to assessors



Persian Task

- ❑ For the first time, a non-European language target collection is part of the CLEF corpus
- ❑ **Persian** is an Indo-European language, spoken in Iran, Afghanistan and Tajikistan, known as **Farsi**.
 - the *Academy of Persian Language and Literature* has declared the name “Persian” is more appropriate than “Farsi”
 - Persian uses **challenging script**, which is a modified version of the Arabic alphabet with elision of short vowels and is written from right to left
 - Persian **morphology is complex** and makes extensive use of suffixes and compounding
- ❑ The task has been organized together with the **Data Base Research Group (DBRG)** of the **University of Tehran** which provided the Hamshahri corpus
- ❑ Both **monolingual** and **bilingual** tasks have been offered



Persian Collection

- ❑ The Hamshahri corpus is a newspaper corpus with news articles from 1996 to 2002, made available by the DBRG of University of Teheran (<http://ece.ut.ac.ir/dbrg/hamshahri/>)
- ❑ News article are categorized both in Persian and English
- ❑ It consists of:
 - size: 628,471,252 bytes
 - items: 166,774 documents

<DOC>

<DOCID>H-750405-266S1</DOCID>

<DOCNO>H-750405-266S1</DOCNO>

<DATE>1996-06-25</DATE>

<CAT xml:lang="fa"> ادب و هنر </CAT>

<CAT xml:lang="en">Literature and Art</CAT>

<TEXT> معرفی کتاب جدید مارکز گابریل گارسیا مارکز، نویسنده کلمبیائیه و برنده جایزه نوبل ادبیات، روز

14 کتاب جدید خود تحت عنوان شرح حال يك آدم ربايه را در مادريد معرفي كرد. كتاب

مذكور شرح حال خانم مارگا پاكن

(Pachon Marga)

و همسرش آلبرتو ویلامیزا

(Villamiza Alberto)

است .

خانم پاكن توسط قاچاقچيان كلمبيا ربه شده بود و همسرش در طول شش ماه براي آزادي وي مبارزه كرد و

شخصا با پابلو اسكوبا (رئيس معروف قاچاقچيان كلمبيائيه) مذاكره كرد. گارسيا مارکز درروز اول سفر خود

به مادريد با فيليپ گونزالس (دوست شخصي اوو نخست وزير سابق اسپانيا) ملاقات كرد

. </TEXT>

</DOC>



Persian Topics

```
<?xml version="1.0" encoding="UTF-8" standalone="no"?>
<topic>
  <identifier>10.2452/599-AH</identifier>
  <title lang="en">2nd of Khordad election</title>
  <title lang="fa">انتخابات دوم خرداد</title>

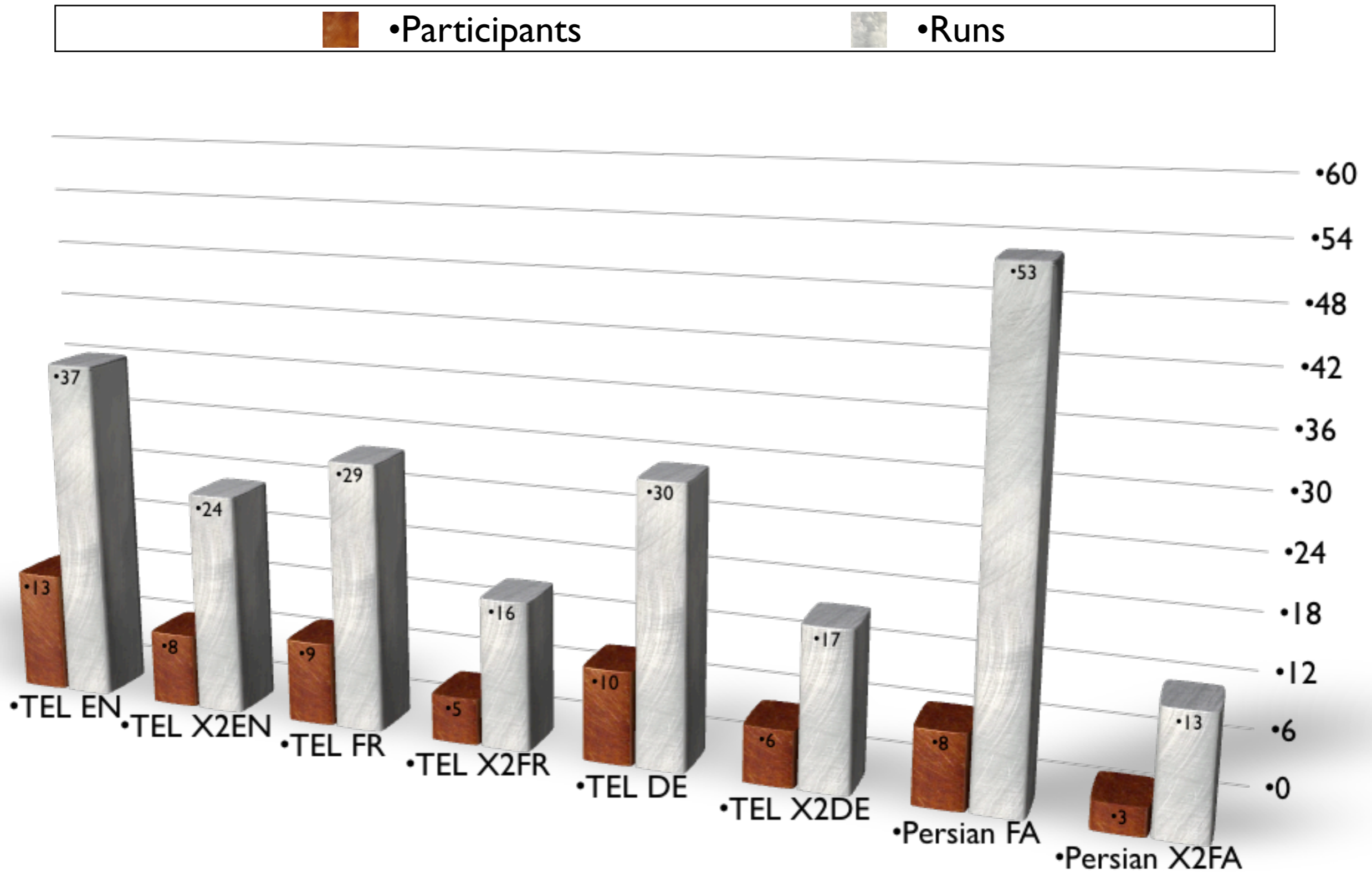
  <description lang="en">Find documents that include information about the 2nd of Khordad
    presidential elections.</description>
  <description lang="fa">سندهایی را پیدا کن که شامل اطلاعاتی در مورد انتخابات دوم خرداد ماه سال 76
    هستند</description>

  <narrative lang="en">Any information about candidates and their sayings, Khatami's unexpected
    winning in the 2nd of Khordad 1376 presidential election is relevant.</narrative>
  <narrative lang="fa">سندهاي مربوط شامل اطلاعاتی در مورد نامزدها و گفته هاي آنها، پیروزي
    غیرمنتظره خاتمي در انتخابات ریاست جمهوری در دوم خرداد ماه سال 76 است</narrative>
</topic>
```

- ❑ 50 topics have been developed in Persian and translated to English
- ❑ Topics consist of **title**, **description**, and **narrative**
- ❑ When translating topics, the attempt is to render them as naturally as possible. This was a particularly difficult task when going from Persian to English as cultural differences had to be catered for

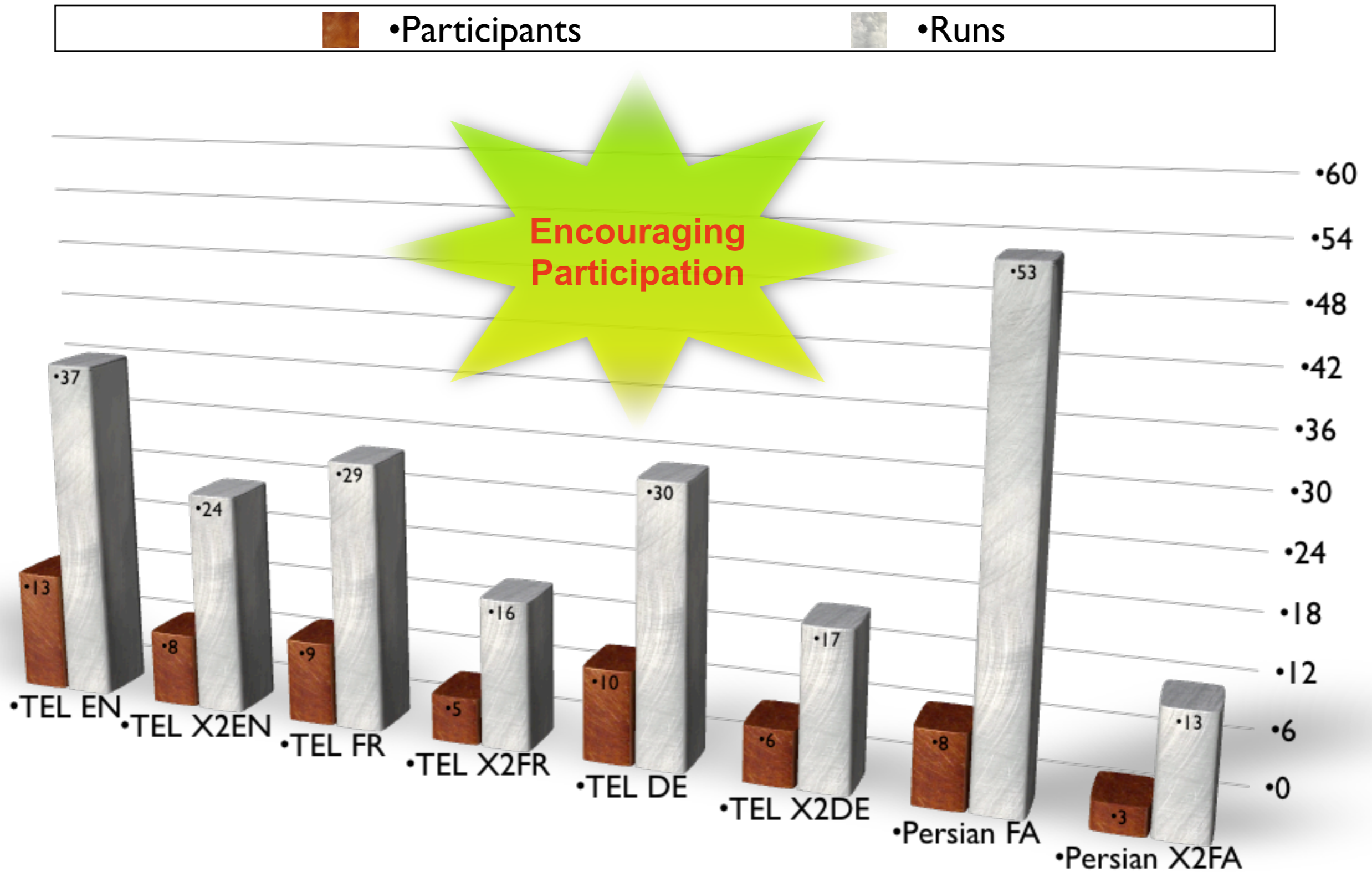


Participation by Task



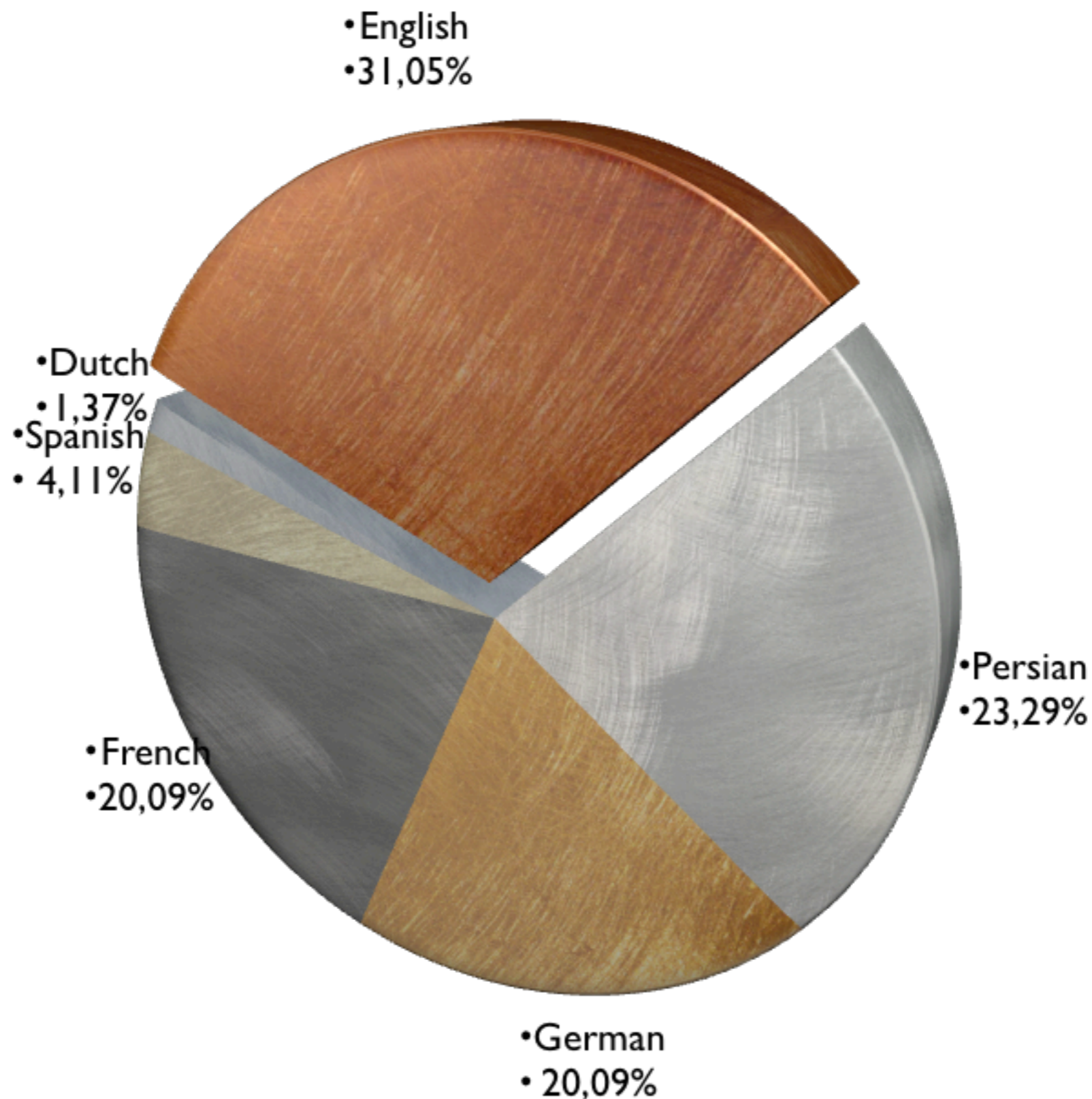


Participation by Task





Experiments by Source Language





Pool Creation

- ❑ Adopted creation criteria:
 - Favour diversity among approaches adopted by participants, according to the descriptions of the experiments provided by the participants;
 - Choose at least one experiment for each participant in each task, chosen among the experiments with highest priority as indicated by the participant;
 - Add mandatory title+description experiments, even though they do not have high priority;
 - Add manual experiments, when provided;
 - For bilingual tasks, ensure that each source topic language is represented.

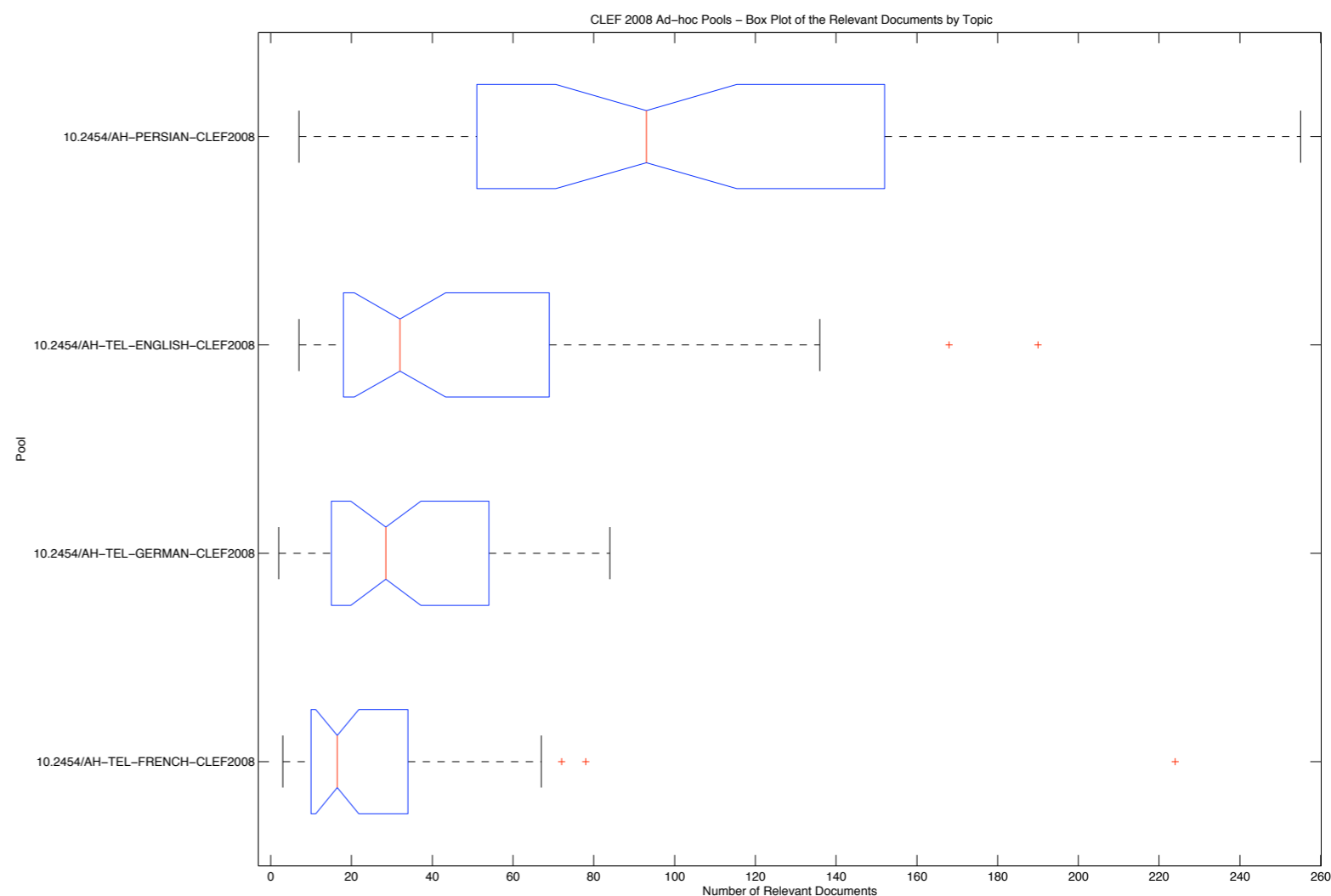
- ❑ Pool depth: 60 documents for all the tasks in order to ensure comparability in the size of the pools

- ❑ Documents taken from the topic creation phase have been added (TEL English and French pools only)



Pool Statistics

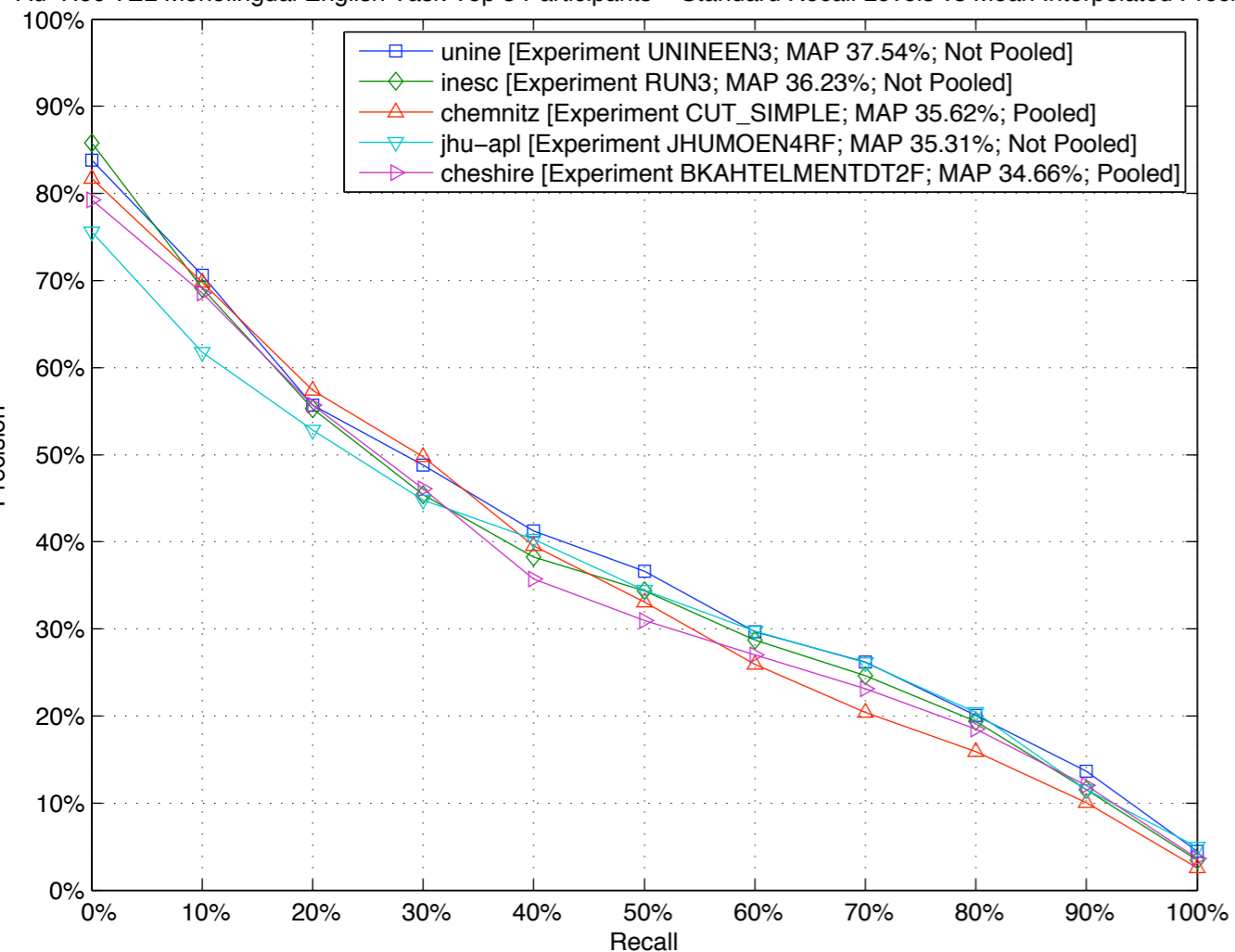
TEL English Pool (DOI 10.2454/AH-TEL-ENGLISH-CLEF2008)	
Pool size	28,104 pooled documents – 25,571 not relevant documents – 2,533 relevant documents 50 topics
Pooled Experiments	21 out of 61 submitted experiments – monolingual: 13 out of 37 submitted experiments – bilingual: 8 out of 24 submitted experiments
Assessors	3 assessors
TEL French Pool (DOI 10.2454/AH-TEL-FRENCH-CLEF2008)	
Pool size	24,530 pooled documents – 23,191 not relevant documents – 1,339 relevant documents 50 topics
Pooled Experiments	14 out of 45 submitted experiments – monolingual: 9 out of 29 submitted experiments – bilingual: 5 out of 16 submitted experiments
Assessors	3 assessors
TEL German Pool (DOI 10.2454/AH-TEL-GERMAN-CLEF2008)	
Pool size	28,734 pooled documents – 27,097 not relevant documents – 1,637 relevant documents 50 topics
Pooled Experiments	16 out of 47 submitted experiments – monolingual: 10 out of 30 submitted experiments – bilingual: 6 out of 17 submitted experiments
Assessors	4 assessors
Persian Pool (DOI 10.2454/AH-PERSIAN-CLEF2008)	
Pool size	26,814 pooled documents – 21,653 not relevant documents – 5,161 relevant documents 50 topics
Pooled Experiments	66 out of 66 submitted experiments – monolingual: 53 out of 53 submitted experiments – bilingual: 13 out of 13 submitted experiments
Assessors	22 assessors



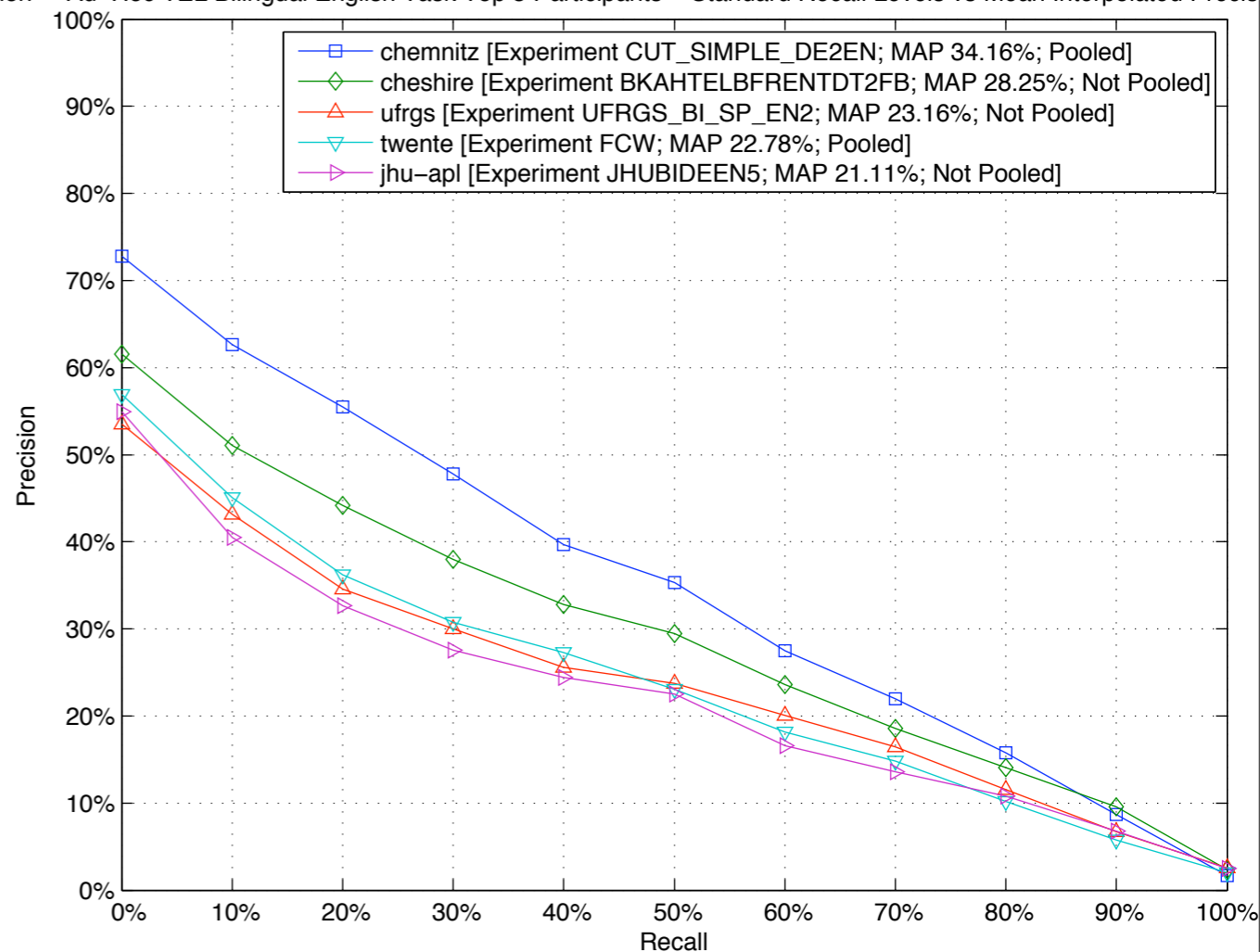


TEL English

Ad-Hoc TEL Monolingual English Task Top 5 Participants – Standard Recall Levels vs Mean Interpolated Precision



Ad-Hoc TEL Bilingual English Task Top 5 Participants – Standard Recall Levels vs Mean Interpolated Precision

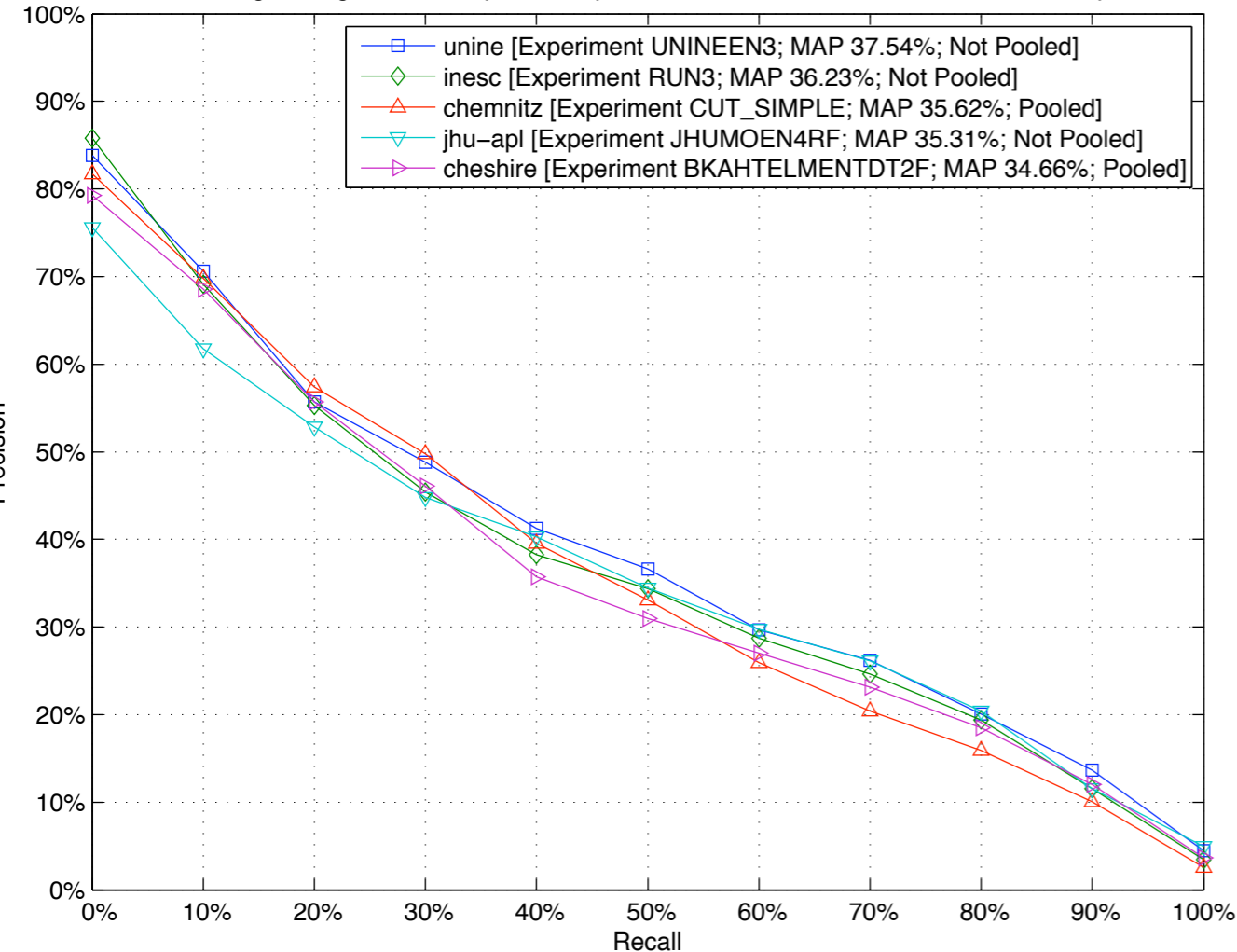




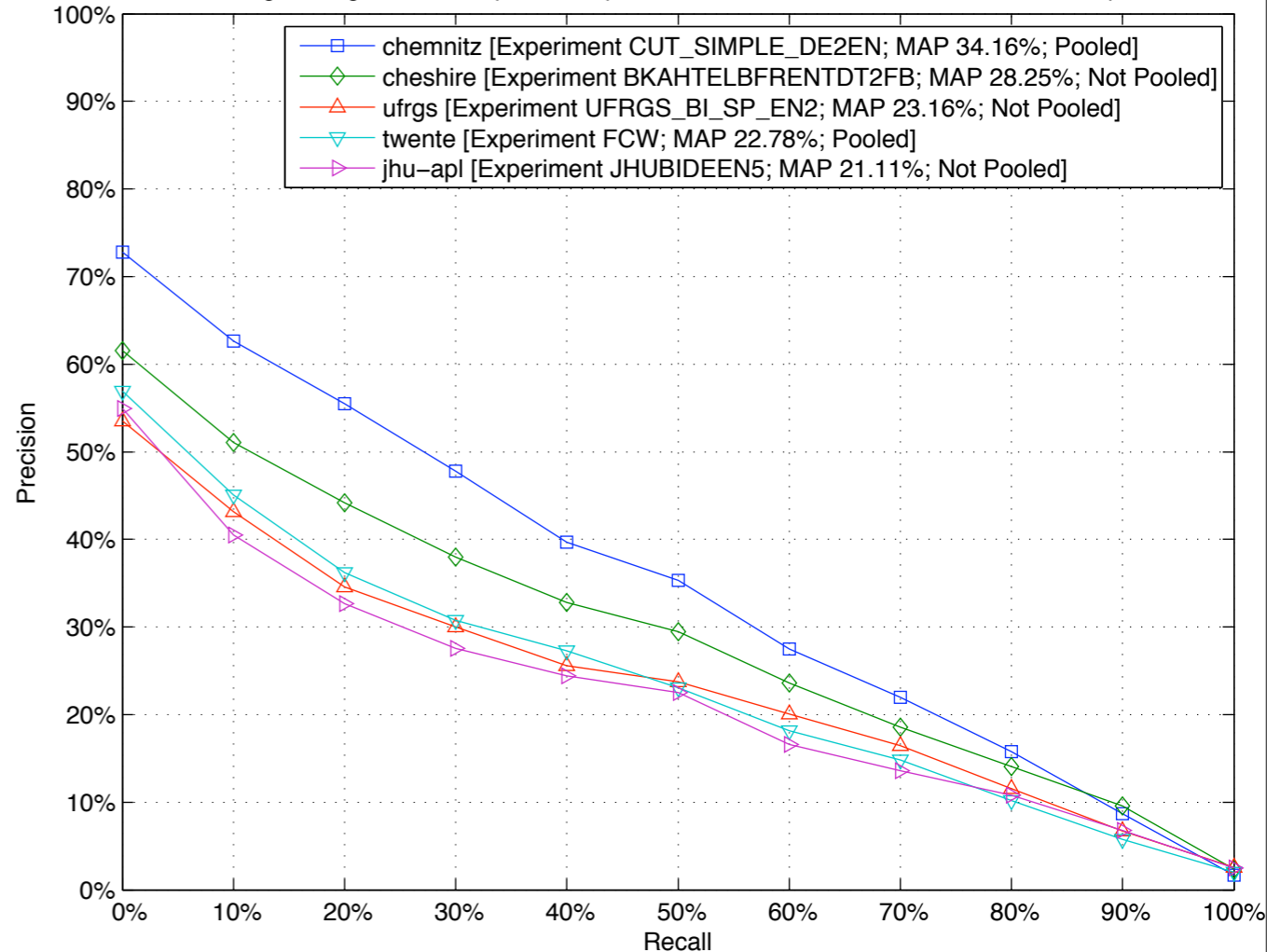
TEL English

**Bilingual is
91%
of monolingual**

Ad-Hoc TEL Monolingual English Task Top 5 Participants – Standard Recall Levels vs Mean Interpolated Precision



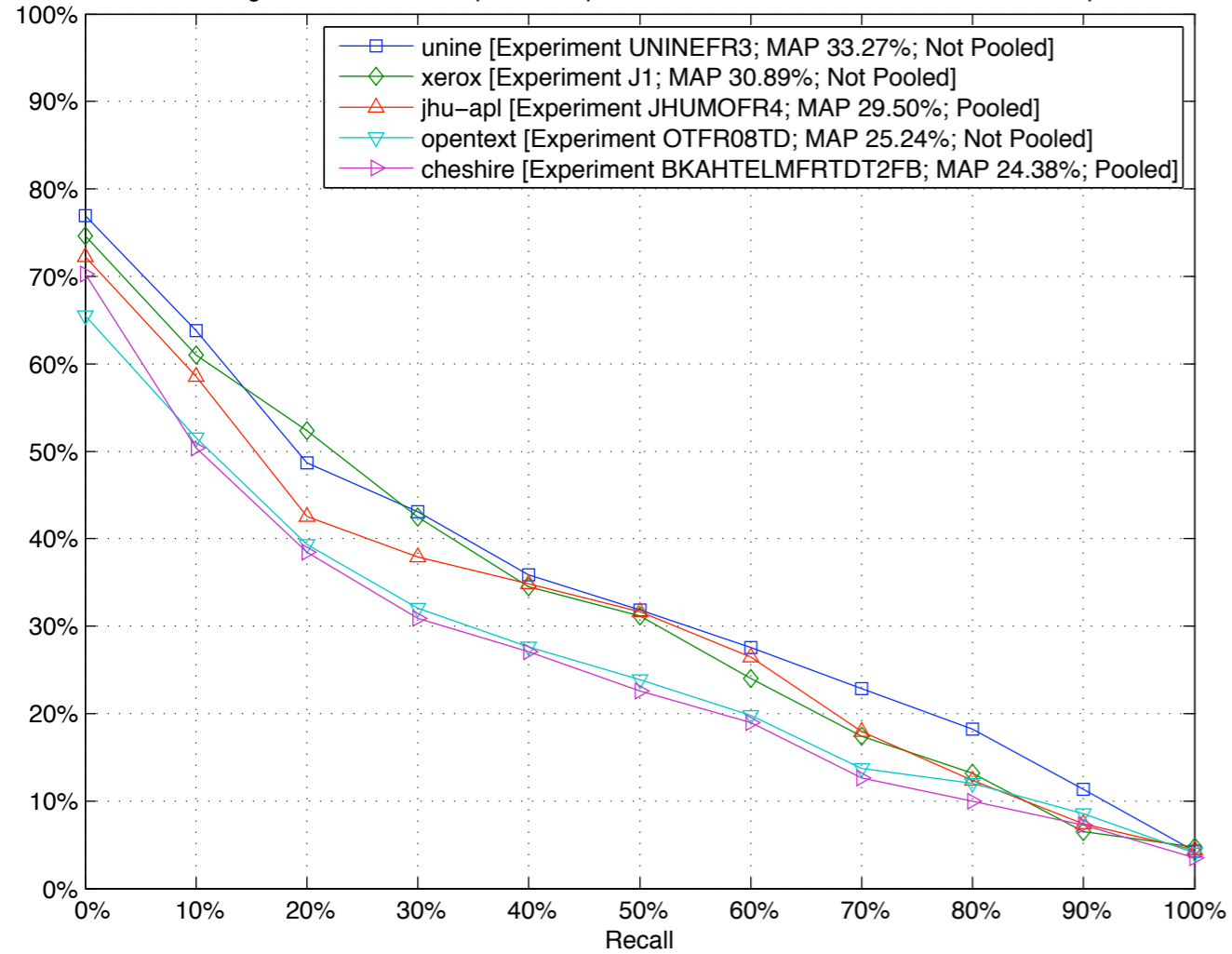
Ad-Hoc TEL Bilingual English Task Top 5 Participants – Standard Recall Levels vs Mean Interpolated Precision



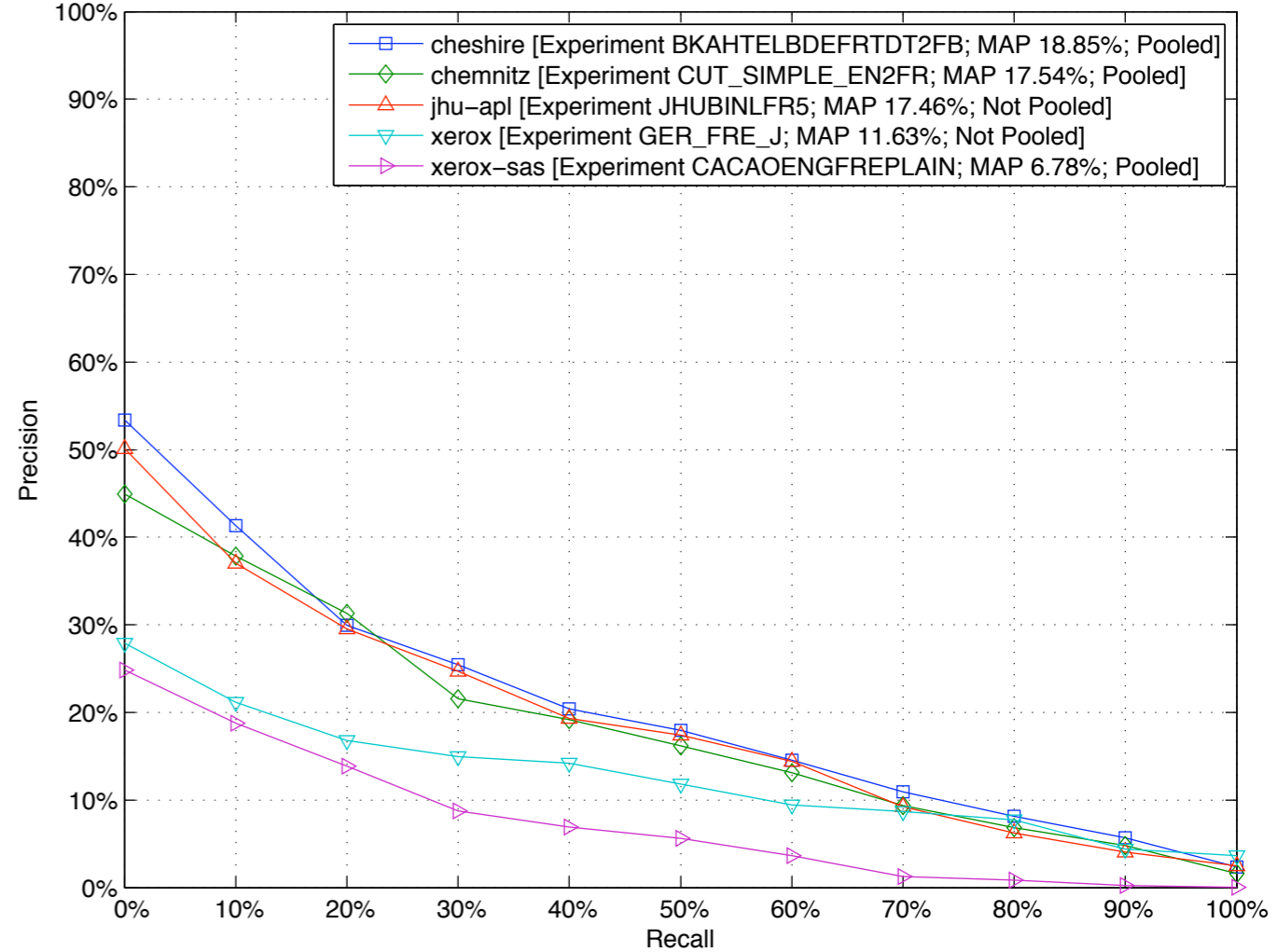


TEL French

Ad-Hoc TEL Monolingual French Task Top 5 Participants – Standard Recall Levels vs Mean Interpolated Precision



Ad-Hoc TEL Bilingual French Task Top 5 Participants – Standard Recall Levels vs Mean Interpolated Precision

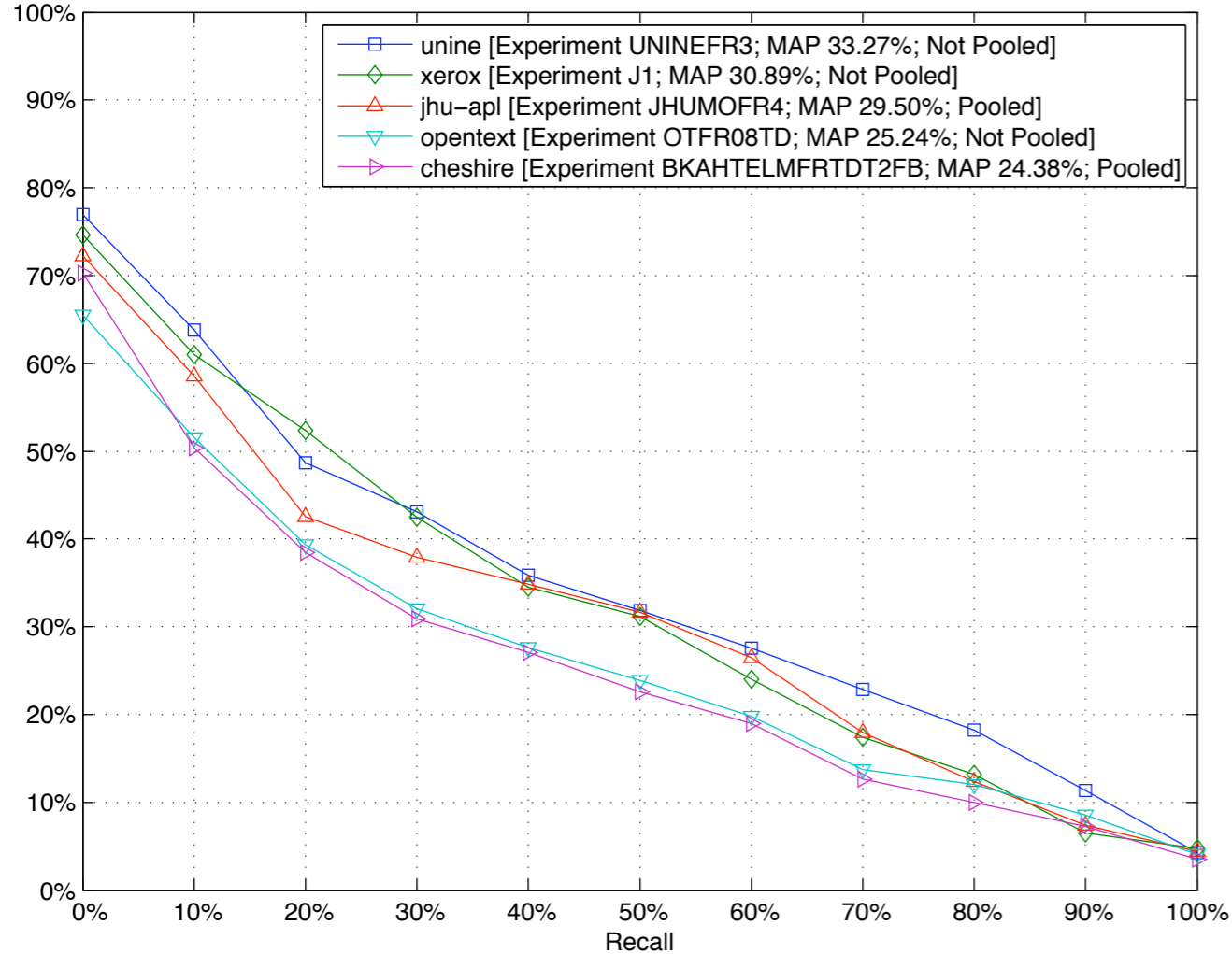




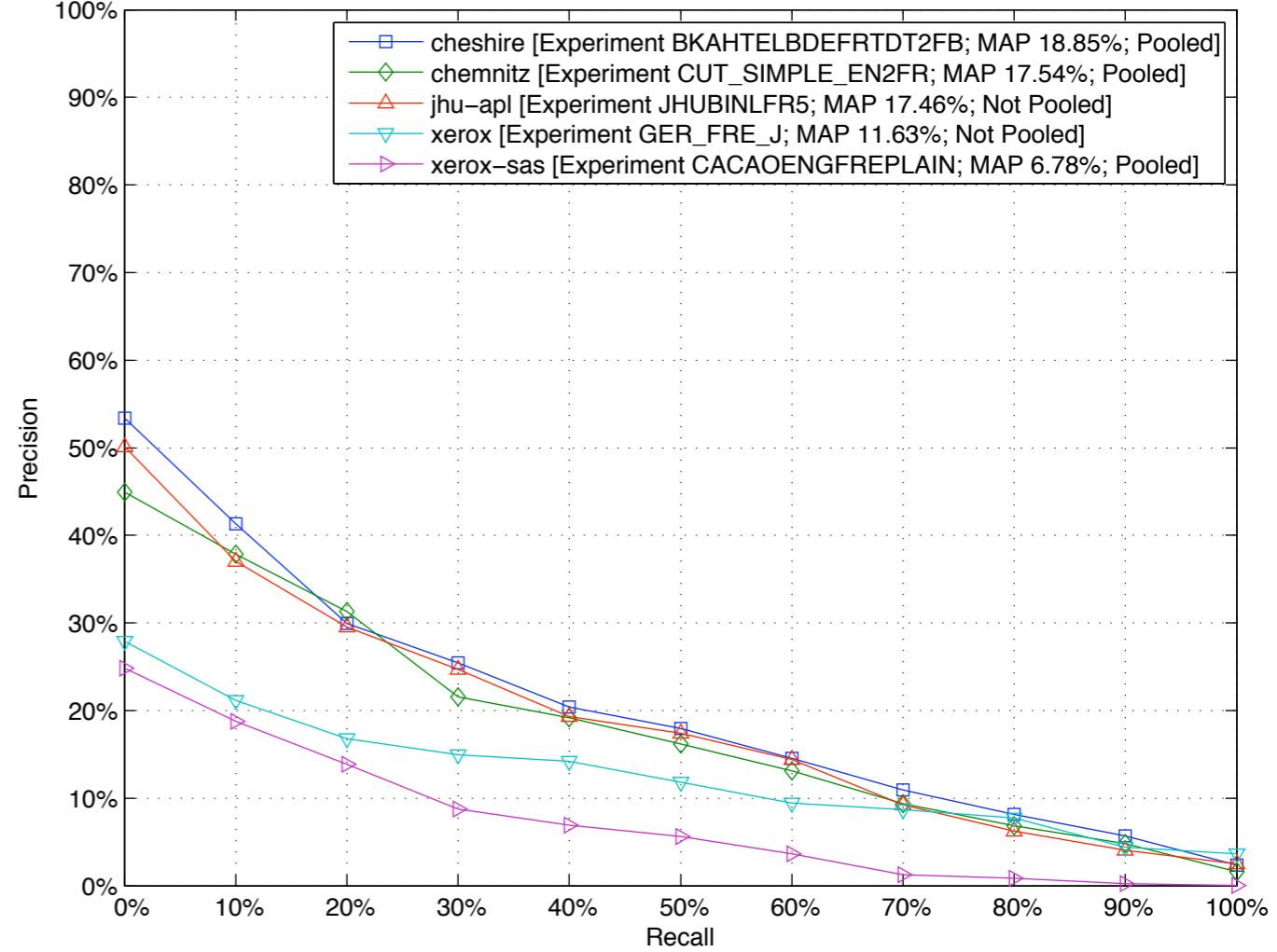
TEL French

**Bilingual is
57%
of monolingual**

Ad-Hoc TEL Monolingual French Task Top 5 Participants – Standard Recall Levels vs Mean Interpolated Precision



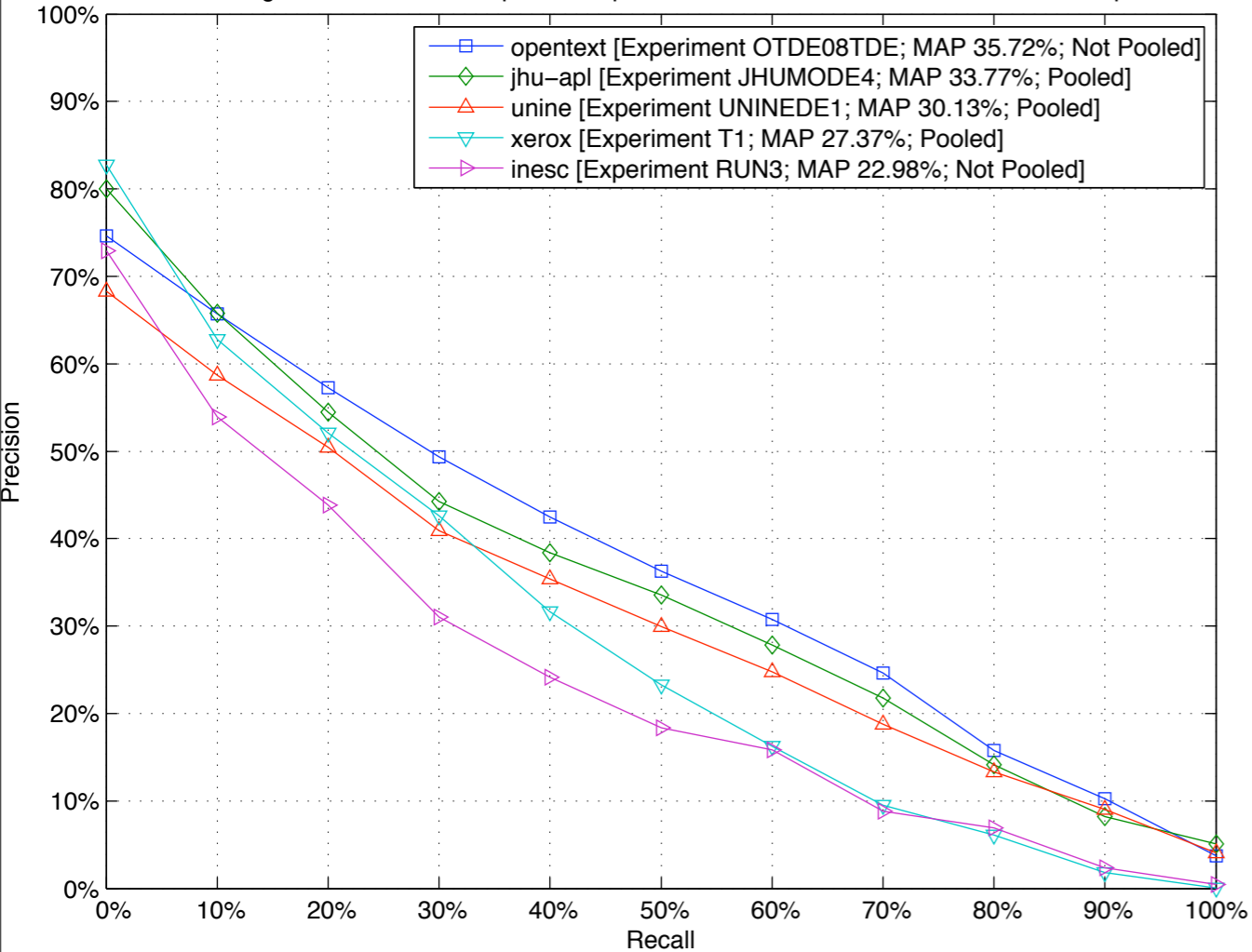
Ad-Hoc TEL Bilingual French Task Top 5 Participants – Standard Recall Levels vs Mean Interpolated Precision



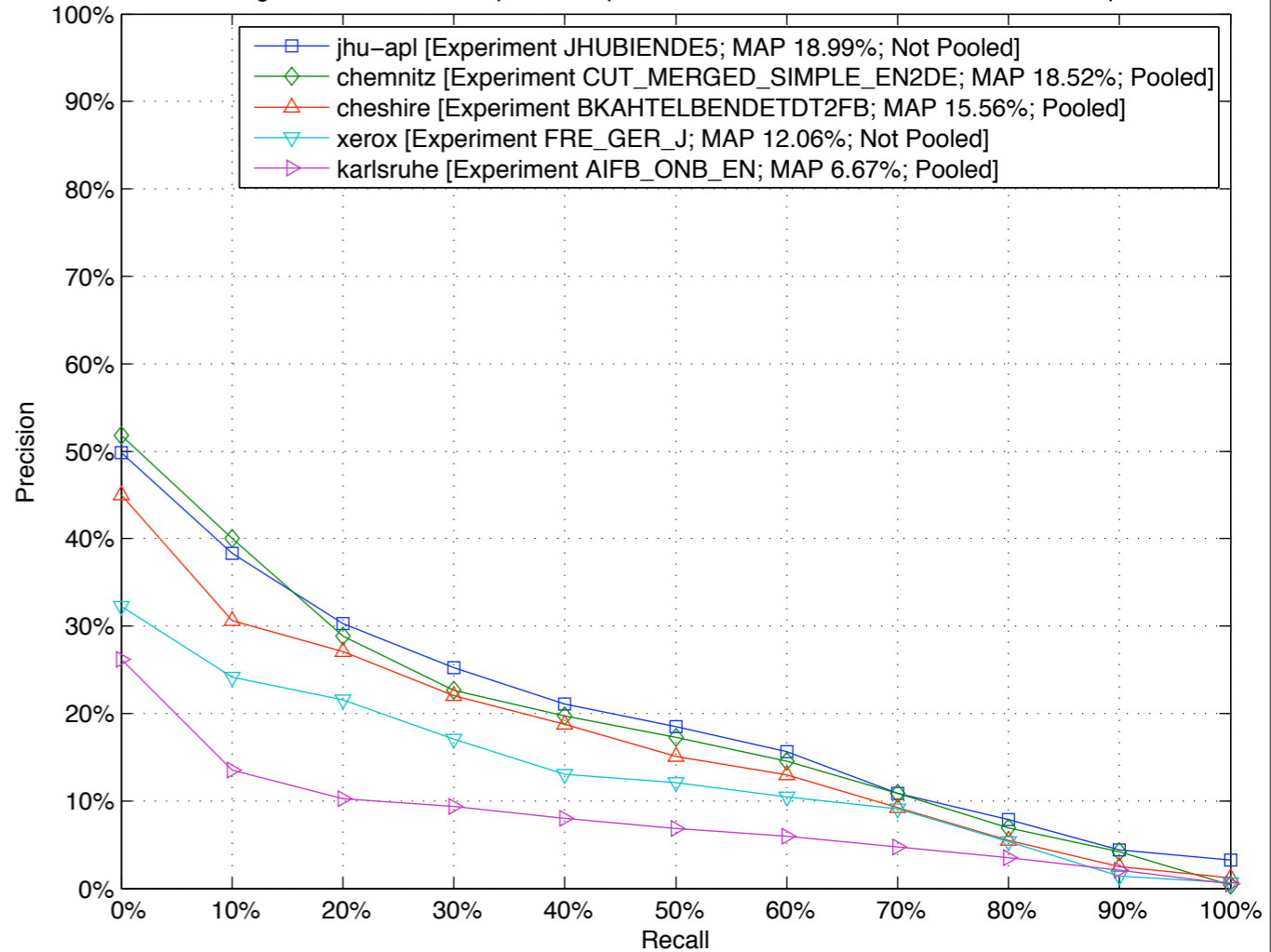


TEL German

Ad-Hoc TEL Monolingual German Task Top 5 Participants – Standard Recall Levels vs Mean Interpolated Precision



Ad-Hoc TEL Bilingual German Task Top 5 Participants – Standard Recall Levels vs Mean Interpolated Precision

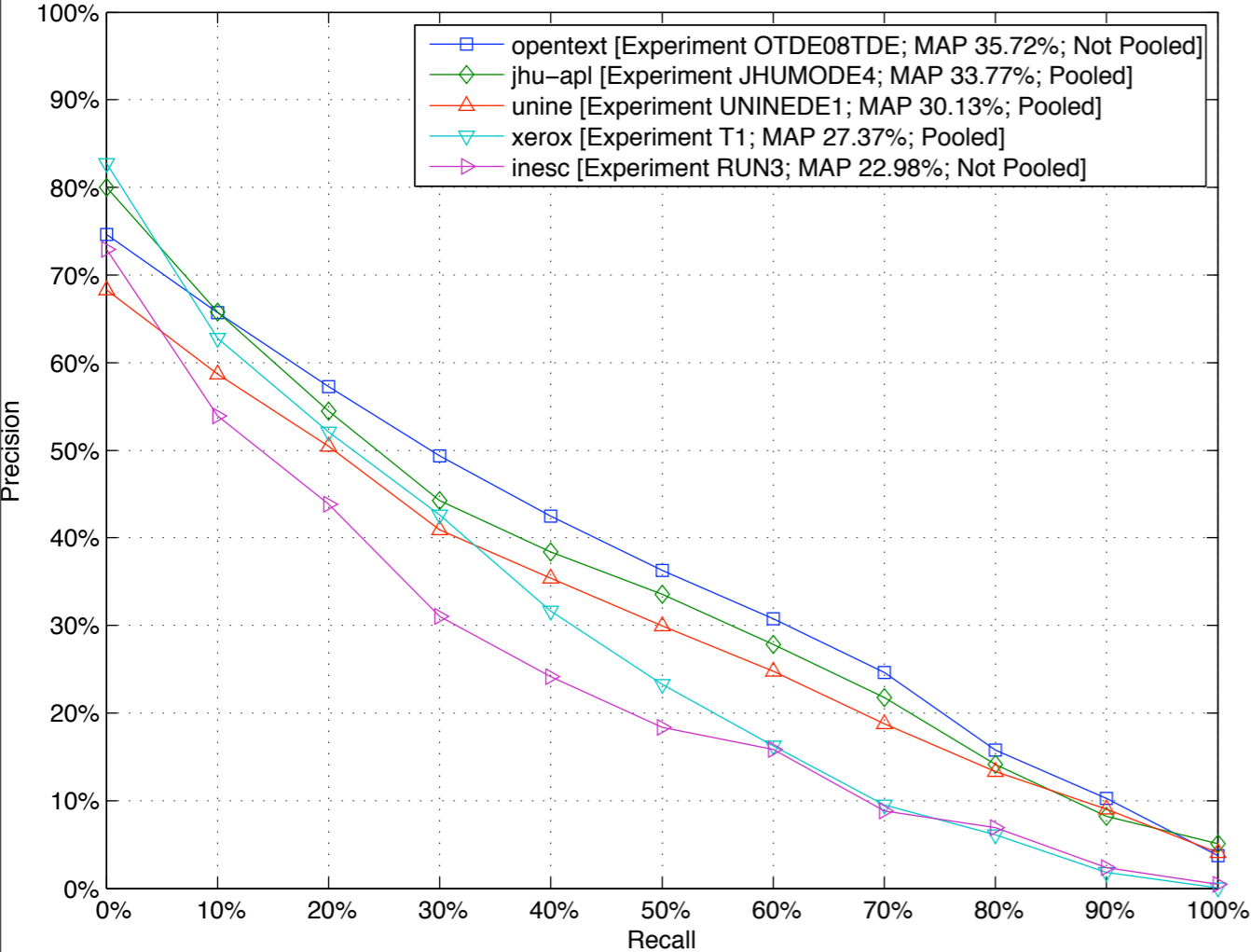




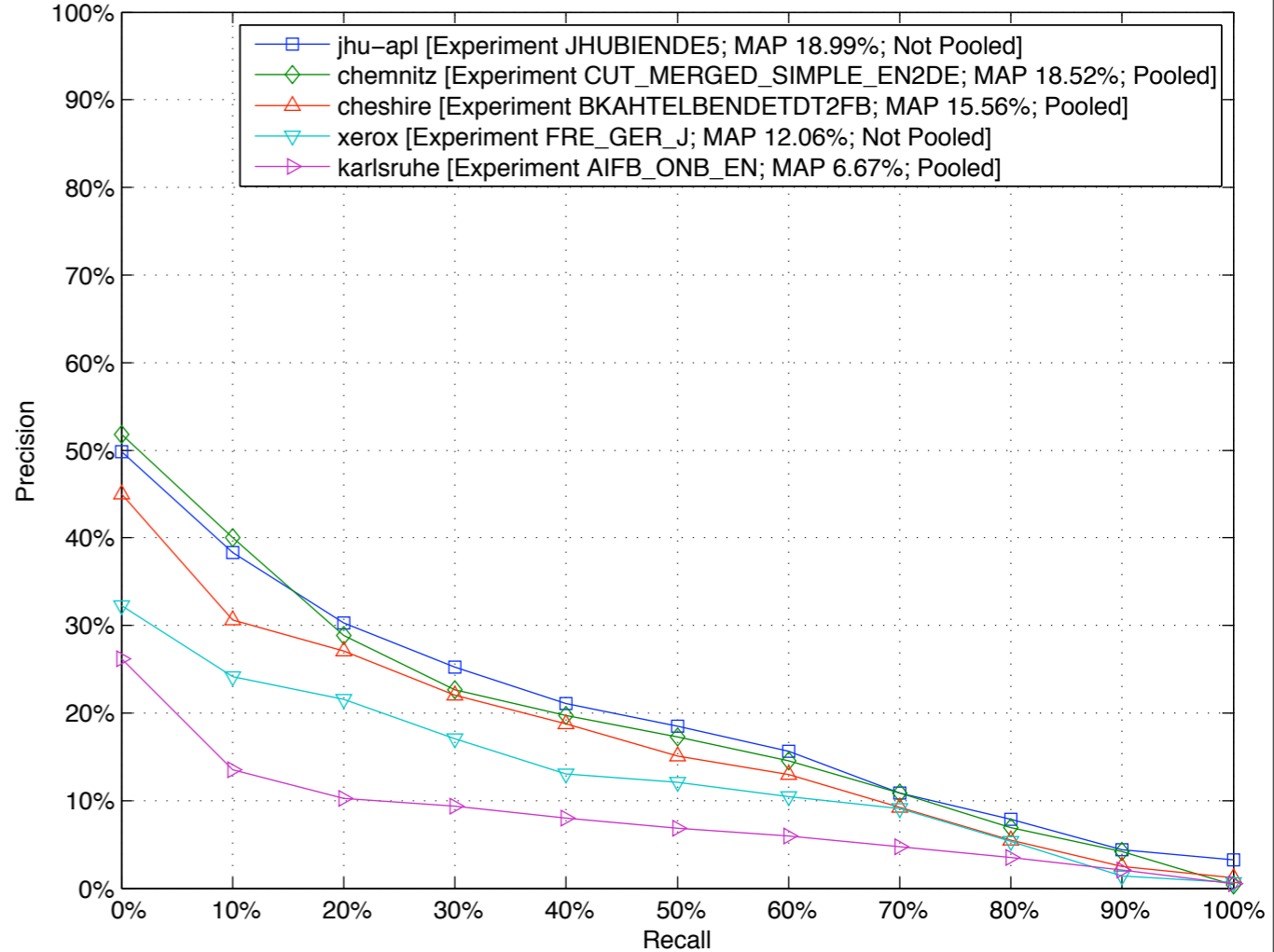
TEL German

**Bilingual is
53%
of monolingual**

Ad-Hoc TEL Monolingual German Task Top 5 Participants – Standard Recall Levels vs Mean Interpolated Precision



Ad-Hoc TEL Bilingual German Task Top 5 Participants – Standard Recall Levels vs Mean Interpolated Precision





TEL: Approaches to Monolingual IR

Track	Rank	Participant	Experiment DOI	MAP	P@5	P@10	P@30
English	1st	unine	10.2415/AH-TEL-MONO-EN-CLEF2008.UNINE.UNINEEN3	37.53%	59.60%	48.66%	39.20%
	2nd	inesc	10.2415/AH-TEL-MONO-EN-CLEF2008.INESC.RUN3	36.23%	61.60%	46.66%	38.00%
	3rd	chemnitz	10.2415/AH-TEL-MONO-EN-CLEF2008.CHEMNITZ.CUT_SIMPLE	35.61%	61.20%	48.53%	38.66%
	4th	jhu-apl	10.2415/AH-TEL-MONO-EN-CLEF2008.JHU-APL.JHUMOEN4RF	35.31%	56.00%	46.66%	39.93%
	5th	cheshire	10.2415/AH-TEL-MONO-EN-CLEF2008.CHESHIRE.BKAHTELMENDT2F	34.66%	60.00%	47.60%	38.33%
	Difference				8.28%	10.00%	4.28%
French	1st	unine	10.2415/AH-TEL-MONO-FR-CLEF2008.UNINE.UNINEFR3	33.27%	48.40%	40.80%	27.53%
	2nd	xerox	10.2415/AH-TEL-MONO-FR-CLEF2008.XEROX.J1	30.88%	46.40%	39.80%	25.86%
	3rd	jhu-apl	10.2415/AH-TEL-MONO-FR-CLEF2008.JHU-APL.JHUMOFR4	29.50%	43.60%	36.20%	23.86%
	4th	opentext	10.2415/AH-TEL-MONO-FR-CLEF2008.OPENTEXT.OTFR08TD	25.23%	36.40%	30.80%	21.86%
	5th	cheshire	10.2415/AH-TEL-MONO-FR-CLEF2008.CHESHIRE.BKAHTELMFRDT2FB	24.37%	37.20%	28.80%	19.46%
	Difference				36.52%	32.96%	41.66%
German	1st	opentext	10.2415/AH-TEL-MONO-DE-CLEF2008.OPENTEXT.OTDE08TDE	35.71%	57.60%	49.60%	33.33%
	2nd	jhu-apl	10.2415/AH-TEL-MONO-DE-CLEF2008.JHU-APL.JHUMODE4	33.77%	58.80%	49.00%	32.60%
	3rd	unine	10.2415/AH-TEL-MONO-DE-CLEF2008.UNINE.UNINEDE1	30.12%	49.60%	44.80%	32.86%
	4th	xerox	10.2415/AH-TEL-MONO-DE-CLEF2008.XEROX.T1	27.36%	53.60%	43.60%	30.80%
	5th	inesc	10.2415/AH-TEL-MONO-DE-CLEF2008.INESC.RUN3	22.97%	48.00%	40.40%	25.40%
	Difference				55.46%	20.00%	22.77%



TEL: Approaches to Monolingual IR

- Traditional IR approach
- Multilinguality of the collection not considered
- Stop-words (around 500 terms)
- Light stemmers + compounding for German
- Probabilistic (Okapi and DFR) and Language Models
- Data fusion with Z-score
- Blind query expansion (Rocchio and idf, 5-10 documents, 10-50 terms)

Track	Rank	Participant	Experiment DOI	MAP	P@5	P@10	P@30
English	1st	unine	10.2415/AH-TEL-MONO-EN-CLEF2008.UNINE.UNINEEN3	37.53%	59.60%	48.66%	39.20%
	2nd	inesc	10.2415/AH-TEL-MONO-EN-CLEF2008.INESC.RUN3	36.23%	61.60%	46.66%	38.00%
	3rd	chemnitz	10.2415/AH-TEL-MONO-EN-CLEF2008.CHEMNITZ.CUT_SIMPLE	35.61%	61.20%	48.53%	38.66%
	4th	jhu-apl	10.2415/AH-TEL-MONO-EN-CLEF2008.JHU-APL.JHUMOEN4RF	35.31%	56.00%	46.66%	39.93%
	5th	cheshire	10.2415/AH-TEL-MONO-EN-CLEF2008.CHESHIRE.BKAHTELMENDDT2F	34.66%	60.00%	47.60%	38.33%
	Difference				8.28%	10.00%	4.28%
French	1st	unine	10.2415/AH-TEL-MONO-FR-CLEF2008.UNINE.UNINEFR3	33.27%	48.40%	40.80%	27.53%
	2nd	xerox	10.2415/AH-TEL-MONO-FR-CLEF2008.XEROX.J1	30.88%	46.40%	39.80%	25.86%
	3rd	jhu-apl	10.2415/AH-TEL-MONO-FR-CLEF2008.JHU-APL.JHUMOFR4	29.50%	43.60%	36.20%	23.86%
	4th	opentext	10.2415/AH-TEL-MONO-FR-CLEF2008.OPENTEXT.OTFR08TD	25.23%	36.40%	30.80%	21.86%
	5th	cheshire	10.2415/AH-TEL-MONO-FR-CLEF2008.CHESHIRE.BKAHTELMFRD2FB	24.37%	37.20%	28.80%	19.46%
	Difference				36.52%	32.96%	41.66%
German	1st	opentext	10.2415/AH-TEL-MONO-DE-CLEF2008.OPENTEXT.OTDE08TDE	35.71%	57.60%	49.60%	33.33%
	2nd	jhu-apl	10.2415/AH-TEL-MONO-DE-CLEF2008.JHU-APL.JHUMODE4	33.77%	58.80%	49.00%	32.60%
	3rd	unine	10.2415/AH-TEL-MONO-DE-CLEF2008.UNINE.UNINEDE1	30.12%	49.60%	44.80%	32.86%
	4th	xerox	10.2415/AH-TEL-MONO-DE-CLEF2008.XEROX.T1	27.36%	53.60%	43.60%	30.80%
	5th	inesc	10.2415/AH-TEL-MONO-DE-CLEF2008.INESC.RUN3	22.97%	48.00%	40.40%	25.40%
	Difference				55.46%	20.00%	22.77%



TEL: Approaches to Monolingual IR

- Traditional IR approach
- Multilinguality of the collection not considered
- Stop-words (few terms)
- Lexicon-based inflectional stemmer and decompounding for German
- Probabilistic (Okapi-like) Model
- Blind query expansion for German

Track	Rank	Participant	Experiment DOI	MAP	P@5	P@10	P@30
English	1st	unine	10.2415/AH-TEL-MONO-EN-CLEF2008.UNINE.UNINEEN3	37.53%	59.60%	48.66%	39.20%
	2nd	inesc	10.2415/AH-TEL-MONO-EN-CLEF2008.INESC.RUN3	36.23%	61.60%	46.66%	38.00%
	3rd	chemnitz	10.2415/AH-TEL-MONO-EN-CLEF2008.CHEMNITZ.CUT_SIMPLE	35.61%	61.20%	48.53%	38.66%
	4th	jhu-apl	10.2415/AH-TEL-MONO-EN-CLEF2008.JHU-APL.JHUMOEN4RF	35.31%	56.00%	46.66%	39.93%
	5th	cheshire	10.2415/AH-TEL-MONO-EN-CLEF2008.CHESHIRE.BKAHTELMENDDT2F	34.66%	60.00%	47.60%	38.33%
	Difference				8.28%	10.00%	4.28%
French	1st	unine	10.2415/AH-TEL-MONO-FR-CLEF2008.UNINE.UNINEFR3	33.27%	48.40%	40.80%	27.53%
	2nd	xerox	10.2415/AH-TEL-MONO-FR-CLEF2008.XEROX.J1	30.88%	46.40%	39.80%	25.86%
	3rd	jhu-apl	10.2415/AH-TEL-MONO-FR-CLEF2008.JHU-APL.JHUMOFR4	29.50%	43.60%	36.20%	23.86%
	4th	opentext	10.2415/AH-TEL-MONO-FR-CLEF2008.OPENTEXT.OTFR08TD	25.23%	36.40%	30.80%	21.86%
	5th	cheshire	10.2415/AH-TEL-MONO-FR-CLEF2008.CHESHIRE.BKAHTELMFRD2FB	24.37%	37.20%	28.80%	19.46%
	Difference				36.52%	32.96%	41.66%
German	1st	opentext	10.2415/AH-TEL-MONO-DE-CLEF2008.OPENTEXT.OTDE08TDE	35.71%	57.60%	49.60%	33.33%
	2nd	jhu-apl	10.2415/AH-TEL-MONO-DE-CLEF2008.JHU-APL.JHUMODE4	33.77%	58.80%	49.00%	32.60%
	3rd	unine	10.2415/AH-TEL-MONO-DE-CLEF2008.UNINE.UNINEDE1	30.12%	49.60%	44.80%	32.86%
	4th	xerox	10.2415/AH-TEL-MONO-DE-CLEF2008.XEROX.T1	27.36%	53.60%	43.60%	30.80%
	5th	inesc	10.2415/AH-TEL-MONO-DE-CLEF2008.INESC.RUN3	22.97%	48.00%	40.40%	25.40%
	Difference				55.46%	20.00%	22.77%



TEL: Approaches to Monolingual IR

- Traditional IR approach but title field weighted three times and description field weighted two times
- Multilinguality of the collection not considered
- Stop-words (120-230 terms)
- Snowball stemmers
- Multinomial language models (Lucene extension made by University of Amsterdam)
- Blind query expansion (Rocchio, 10 documents, 200 terms, LucQE extension)

Track	Rank	Participant	Experiment DOI	MAP	P@5	P@10	P@30
English	1st	unine	10.2415/AH-TEL-MONO-EN-CLEF2008.UNINE.UNINEEN3	37.53%	59.60%	48.66%	39.20%
	2nd	inesc	10.2415/AH-TEL-MONO-EN-CLEF2008.INESC.RUN3	36.23%	61.60%	46.66%	38.00%
	3rd	chemnitz	10.2415/AH-TEL-MONO-EN-CLEF2008.CHEMNITZ.CUT_SIMPLE	35.61%	61.20%	48.53%	38.66%
	4th	jhu-apl	10.2415/AH-TEL-MONO-EN-CLEF2008.JHU-APL.JHUMOEN4RF	35.31%	56.00%	46.66%	39.93%
	5th	cheshire	10.2415/AH-TEL-MONO-EN-CLEF2008.CHESHIRE.BKAHTELMENDT2F	34.66%	60.00%	47.60%	38.33%
	Difference				8.28%	10.00%	4.28%
French	1st	unine	10.2415/AH-TEL-MONO-FR-CLEF2008.UNINE.UNINEFR3	33.27%	48.40%	40.80%	27.53%
	2nd	xerox	10.2415/AH-TEL-MONO-FR-CLEF2008.XEROX.J1	30.88%	46.40%	39.80%	25.86%
	3rd	jhu-apl	10.2415/AH-TEL-MONO-FR-CLEF2008.JHU-APL.JHUMOFR4	29.50%	43.60%	36.20%	23.86%
	4th	opentext	10.2415/AH-TEL-MONO-FR-CLEF2008.OPENTEXT.OTFR08TD	25.23%	36.40%	30.80%	21.86%
	5th	cheshire	10.2415/AH-TEL-MONO-FR-CLEF2008.CHESHIRE.BKAHTELMFRDT2FB	24.37%	37.20%	28.80%	19.46%
	Difference				36.52%	32.96%	41.66%
German	1st	opentext	10.2415/AH-TEL-MONO-DE-CLEF2008.OPENTEXT.OTDE08TDE	35.71%	57.60%	49.60%	33.33%
	2nd	jhu-apl	10.2415/AH-TEL-MONO-DE-CLEF2008.JHU-APL.JHUMODE4	33.77%	58.80%	49.00%	32.60%
	3rd	unine	10.2415/AH-TEL-MONO-DE-CLEF2008.UNINE.UNINEDE1	30.12%	49.60%	44.80%	32.86%
	4th	xerox	10.2415/AH-TEL-MONO-DE-CLEF2008.XEROX.T1	27.36%	53.60%	43.60%	30.80%
	5th	inesc	10.2415/AH-TEL-MONO-DE-CLEF2008.INESC.RUN3	22.97%	48.00%	40.40%	25.40%
	Difference				55.46%	20.00%	22.77%



TEL: Approaches to Monolingual IR

- Traditional IR approach
- Multilinguality of the collection considered: automatic identification of the language (not used due to an error) and translation of the query into top 10 languages (not top run)
- Stop-words (unine)
- Porter & Krovetz for English, Snowball stemmers for French and German plus n-gram decomposer for German
- Vector space Model (Lucene)
- Blind query expansion
- Data fusion (Z-score) when different stemmers have been used

Track	Rank	Participant	Experiment DOI	MAP	P@5	P@10	P@30
English	1st	unine	10.2415/AH-TEL-MONO-EN-CLEF2008.UNINE.UNINEEN3	37.53%	59.60%	48.66%	39.20%
	2nd	inesc	10.2415/AH-TEL-MONO-EN-CLEF2008.INESC.RUN3	36.23%	61.60%	46.66%	38.00%
	3rd	chemnitz	10.2415/AH-TEL-MONO-EN-CLEF2008.CHEMNITZ.CUT_SIMPLE	35.61%	61.20%	48.53%	38.66%
	4th	jhu-apl	10.2415/AH-TEL-MONO-EN-CLEF2008.JHU-APL.JHUMOEN4RF	35.31%	56.00%	46.66%	39.93%
	5th	cheshire	10.2415/AH-TEL-MONO-EN-CLEF2008.CHESHIRE.BKAHTELMENDT2F	34.66%	60.00%	47.60%	38.33%
	Difference				8.28%	10.00%	4.28%
French	1st	unine	10.2415/AH-TEL-MONO-FR-CLEF2008.UNINE.UNINEFR3	33.27%	48.40%	40.80%	27.53%
	2nd	xerox	10.2415/AH-TEL-MONO-FR-CLEF2008.XEROX.J1	30.88%	46.40%	39.80%	25.86%
	3rd	jhu-apl	10.2415/AH-TEL-MONO-FR-CLEF2008.JHU-APL.JHUMOFR4	29.50%	43.60%	36.20%	23.86%
	4th	opentext	10.2415/AH-TEL-MONO-FR-CLEF2008.OPENTEXT.OTFR08TD	25.23%	36.40%	30.80%	21.86%
	5th	cheshire	10.2415/AH-TEL-MONO-FR-CLEF2008.CHESHIRE.BKAHTELMFRDT2FB	24.37%	37.20%	28.80%	19.46%
	Difference				36.52%	32.96%	41.66%
German	1st	opentext	10.2415/AH-TEL-MONO-DE-CLEF2008.OPENTEXT.OTDE08TDE	35.71%	57.60%	49.60%	33.33%
	2nd	jhu-apl	10.2415/AH-TEL-MONO-DE-CLEF2008.JHU-APL.JHUMODE4	33.77%	58.80%	49.00%	32.60%
	3rd	unine	10.2415/AH-TEL-MONO-DE-CLEF2008.UNINE.UNINEDE1	30.12%	49.60%	44.80%	32.86%
	4th	xerox	10.2415/AH-TEL-MONO-DE-CLEF2008.XEROX.T1	27.36%	53.60%	43.60%	30.80%
	5th	inesc	10.2415/AH-TEL-MONO-DE-CLEF2008.INESC.RUN3	22.97%	48.00%	40.40%	25.40%
	Difference				55.46%	20.00%	22.77%



TEL: Approaches to Monolingual IR

- Traditional IR approach
- Multilinguality of the collection considered: creation of a single multilingual index (meta-language) and translate the query to the meta-language, giving more weight to the main language of the collection
- Stop-words
- In-site stemmers and decompounder for German
- Language model (Lemur)

Track	Rank	Participant	Experiment DOI	MAP	P@5	P@10	P@30
English	1st	unine	10.2415/AH-TEL-MONO-EN-CLEF2008.UNINE.UNINEEN3	37.53%	59.60%	48.66%	39.20%
	2nd	inesc	10.2415/AH-TEL-MONO-EN-CLEF2008.INESC.RUN3	36.23%	61.60%	46.66%	38.00%
	3rd	chemnitz	10.2415/AH-TEL-MONO-EN-CLEF2008.CHEMNITZ.CUT_SIMPLE	35.61%	61.20%	48.53%	38.66%
	4th	jhu-apl	10.2415/AH-TEL-MONO-EN-CLEF2008.JHU-APL.JHUMOEN4RF	35.31%	56.00%	46.66%	39.93%
	5th	cheshire	10.2415/AH-TEL-MONO-EN-CLEF2008.CHESHIRE.BKAHTELMEN TDT2F	34.66%	60.00%	47.60%	38.33%
	Difference				8.28%	10.00%	4.28%
French	1st	unine	10.2415/AH-TEL-MONO-FR-CLEF2008.UNINE.UNINEFR3	33.27%	48.40%	40.80%	27.53%
	2nd	xerox	10.2415/AH-TEL-MONO-FR-CLEF2008.XEROX.J1	30.88%	46.40%	39.80%	25.86%
	3rd	jhu-apl	10.2415/AH-TEL-MONO-FR-CLEF2008.JHU-APL.JHUMOFR4	29.50%	43.60%	36.20%	23.86%
	4th	opentext	10.2415/AH-TEL-MONO-FR-CLEF2008.OPENTEXT.OTFR08TD	25.23%	36.40%	30.80%	21.86%
	5th	cheshire	10.2415/AH-TEL-MONO-FR-CLEF2008.CHESHIRE.BKAHTELMFR TDT2FB	24.37%	37.20%	28.80%	19.46%
	Difference				36.52%	32.96%	41.66%
German	1st	opentext	10.2415/AH-TEL-MONO-DE-CLEF2008.OPENTEXT.OTDE08TDE	35.71%	57.60%	49.60%	33.33%
	2nd	jhu-apl	10.2415/AH-TEL-MONO-DE-CLEF2008.JHU-APL.JHUMODE4	33.77%	58.80%	49.00%	32.60%
	3rd	unine	10.2415/AH-TEL-MONO-DE-CLEF2008.UNINE.UNINEDE1	30.12%	49.60%	44.80%	32.86%
	4th	xerox	10.2415/AH-TEL-MONO-DE-CLEF2008.XEROX.T1	27.36%	53.60%	43.60%	30.80%
	5th	inesc	10.2415/AH-TEL-MONO-DE-CLEF2008.INESC.RUN3	22.97%	48.00%	40.40%	25.40%
	Difference				55.46%	20.00%	22.77%



TEL: Approaches to Monolingual IR

- Traditional IR approach
- Multilinguality of the collection not considered
- 4-grams stemmers
- Language model with smoothing between term frequencies and corpus frequencies
- Blind query expansion (only English run in the top 5)

Track	Rank	Participant	Experiment DOI	MAP	P@5	P@10	P@30
English	1st	unine	10.2415/AH-TEL-MONO-EN-CLEF2008.UNINE.UNINEEN3	37.53%	59.60%	48.66%	39.20%
	2nd	inesc	10.2415/AH-TEL-MONO-EN-CLEF2008.INESC.RUN3	36.23%	61.60%	46.66%	38.00%
	3rd	chemnitz	10.2415/AH-TEL-MONO-EN-CLEF2008.CHEMNITZ.CUT_SIMPLE	35.61%	61.20%	48.53%	38.66%
	4th	jhu-apl	10.2415/AH-TEL-MONO-EN-CLEF2008.JHU-APL.JHUMOEN4RF	35.31%	56.00%	46.66%	39.93%
	5th	cheshire	10.2415/AH-TEL-MONO-EN-CLEF2008.CHESHIRE.BKAHTELMENDDT2F	34.66%	60.00%	47.60%	38.33%
	Difference				8.28%	10.00%	4.28%
French	1st	unine	10.2415/AH-TEL-MONO-FR-CLEF2008.UNINE.UNINEFR3	33.27%	48.40%	40.80%	27.53%
	2nd	xerox	10.2415/AH-TEL-MONO-FR-CLEF2008.XEROX.J1	30.88%	46.40%	39.80%	25.86%
	3rd	jhu-apl	10.2415/AH-TEL-MONO-FR-CLEF2008.JHU-APL.JHUMOFR4	29.50%	43.60%	36.20%	23.86%
	4th	opentext	10.2415/AH-TEL-MONO-FR-CLEF2008.OPENTEXT.OTFR08TD	25.23%	36.40%	30.80%	21.86%
	5th	cheshire	10.2415/AH-TEL-MONO-FR-CLEF2008.CHESHIRE.BKAHTELMFRD2FB	24.37%	37.20%	28.80%	19.46%
	Difference				36.52%	32.96%	41.66%
German	1st	opentext	10.2415/AH-TEL-MONO-DE-CLEF2008.OPENTEXT.OTDE08TDE	35.71%	57.60%	49.60%	33.33%
	2nd	jhu-apl	10.2415/AH-TEL-MONO-DE-CLEF2008.JHU-APL.JHUMODE4	33.77%	58.80%	49.00%	32.60%
	3rd	unine	10.2415/AH-TEL-MONO-DE-CLEF2008.UNINE.UNINEDE1	30.12%	49.60%	44.80%	32.86%
	4th	xerox	10.2415/AH-TEL-MONO-DE-CLEF2008.XEROX.T1	27.36%	53.60%	43.60%	30.80%
	5th	inesc	10.2415/AH-TEL-MONO-DE-CLEF2008.INESC.RUN3	22.97%	48.00%	40.40%	25.40%
	Difference				55.46%	20.00%	22.77%



TEL: Approaches to Monolingual IR

- Structure of the record is considered but not fully exploited
- Multilinguality of the collection not considered
- Stop-words
- Snowball stemmers (no decompounding for German)
- Logistic regression
- Blind query expansion (probabilistic RF, 10 documents, 10 terms)

Track	Rank	Participant	Experiment DOI	MAP	P@5	P@10	P@30
English	1st	unine	10.2415/AH-TEL-MONO-EN-CLEF2008.UNINE.UNINEEN3	37.53%	59.60%	48.66%	39.20%
	2nd	inesc	10.2415/AH-TEL-MONO-EN-CLEF2008.INESC.RUN3	36.23%	61.60%	46.66%	38.00%
	3rd	chemnitz	10.2415/AH-TEL-MONO-EN-CLEF2008.CHEMNITZ.CUT_SIMPLE	35.61%	61.20%	48.53%	38.66%
	4th	jhu-apl	10.2415/AH-TEL-MONO-EN-CLEF2008.JHU-APL.JHUMOEN4RF	35.31%	56.00%	46.66%	39.93%
	5th	cheshire	10.2415/AH-TEL-MONO-EN-CLEF2008.CHESHIRE.BKAHTELMEN TDT2F	34.66%	60.00%	47.60%	38.33%
	Difference				8.28%	10.00%	4.28%
French	1st	unine	10.2415/AH-TEL-MONO-FR-CLEF2008.UNINE.UNINEFR3	33.27%	48.40%	40.80%	27.53%
	2nd	xerox	10.2415/AH-TEL-MONO-FR-CLEF2008.XEROX.J1	30.88%	46.40%	39.80%	25.86%
	3rd	jhu-apl	10.2415/AH-TEL-MONO-FR-CLEF2008.JHU-APL.JHUMOFR4	29.50%	43.60%	36.20%	23.86%
	4th	opentext	10.2415/AH-TEL-MONO-FR-CLEF2008.OPENTEXT.OTFR08TD	25.23%	36.40%	30.80%	21.86%
	5th	chesire	10.2415/AH-TEL-MONO-FR-CLEF2008.CHESHIRE.BKAHTELMFR TDT2FB	24.37%	37.20%	28.80%	19.46%
	Difference				36.52%	32.96%	41.66%
German	1st	opentext	10.2415/AH-TEL-MONO-DE-CLEF2008.OPENTEXT.OTDE08TDE	35.71%	57.60%	49.60%	33.33%
	2nd	jhu-apl	10.2415/AH-TEL-MONO-DE-CLEF2008.JHU-APL.JHUMODE4	33.77%	58.80%	49.00%	32.60%
	3rd	unine	10.2415/AH-TEL-MONO-DE-CLEF2008.UNINE.UNINEDE1	30.12%	49.60%	44.80%	32.86%
	4th	xerox	10.2415/AH-TEL-MONO-DE-CLEF2008.XEROX.T1	27.36%	53.60%	43.60%	30.80%
	5th	inesc	10.2415/AH-TEL-MONO-DE-CLEF2008.INESC.RUN3	22.97%	48.00%	40.40%	25.40%
	Difference				55.46%	20.00%	22.77%



TEL: Approaches to Monolingual IR

Track	Rank	Participant	Experiment DOI	MAP	P@5	P@10	P@30
English	1st	unine	10.2415/AH-TEL-MONO-EN-CLEF2008.UNINE.UNINEEN3	37.53%	59.60%	48.66%	39.20%
	2nd	inesc	10.2415/AH-TEL-MONO-EN-CLEF2008.INESC.RUN3	36.23%	61.60%	46.66%	38.00%
	3rd	chemnitz	10.2415/AH-TEL-MONO-EN-CLEF2008.CHEMNITZ.CUT_SIMPLE	35.61%	61.20%	48.53%	38.66%
	4th	jhu-apl	10.2415/AH-TEL-MONO-EN-CLEF2008.JHU-APL.JHUMOEN4RF	35.31%	56.00%	46.66%	39.93%
	5th	cheshire	10.2415/AH-TEL-MONO-EN-CLEF2008.CHESHIRE.BKAHTELMEN TDT2F	34.66%	60.00%	47.60%	38.33%
	Difference				8.28%	10.00%	4.28%
French	1st	unine	10.2415/AH-TEL-MONO-FR-CLEF2008.UNINE.UNINEFR3	33.27%	48.40%	40.80%	27.53%
	2nd	xerox	10.2415/AH-TEL-MONO-FR-CLEF2008.XEROX.J1	30.88%	46.40%	39.80%	25.86%
	3rd	jhu-apl	10.2415/AH-TEL-MONO-FR-CLEF2008.JHU-APL.JHUMOFR4	29.50%	43.60%	36.20%	23.86%
	4th	opentext	10.2415/AH-TEL-MONO-FR-CLEF2008.OPENTEXT.OTFR08TD	25.23%	36.40%	30.80%	21.86%
	5th	chesire	10.2415/AH-TEL-MONO-FR-CLEF2008.CHESHIRE.BKAHTELMFR TDT2FB	24.37%	37.20%	28.80%	19.46%
	Difference				36.52%	32.96%	41.66%
German	1st	opentext	10.2415/AH-TEL-MONO-DE-CLEF2008.OPENTEXT.OTDE08TDE	35.71%	57.60%	49.60%	33.33%
	2nd	jhu-apl	10.2415/AH-TEL-MONO-DE-CLEF2008.JHU-APL.JHUMODE4	33.77%	58.80%	49.00%	32.60%
	3rd	unine	10.2415/AH-TEL-MONO-DE-CLEF2008.UNINE.UNINEDE1	30.12%	49.60%	44.80%	32.86%
	4th	xerox	10.2415/AH-TEL-MONO-DE-CLEF2008.XEROX.T1	27.36%	53.60%	43.60%	30.80%
	5th	inesc	10.2415/AH-TEL-MONO-DE-CLEF2008.INESC.RUN3	22.97%	48.00%	40.40%	25.40%
	Difference				55.46%	20.00%	22.77%



TEL: Approaches to Bilingual IR

Track	Rank	Participant	Experiment DOI	MAP	P@5	P@10	P@30
English	1st	chemnitz	10.2415/AH-TEL-BILI-X2EN-CLEF2008.CHEMNITZ.CUT_SIMPLE_DE2EN	34.15%	55.20%	50.20%	38.73%
	2nd	chesire	10.2415/AH-TEL-BILI-X2EN-CLEF2008.CESHIRE.BKAHTELBFRENTDT2FB	28.24%	46.40%	40.99%	30.73%
	3rd	ufrgs	10.2415/AH-TEL-BILI-X2EN-CLEF2008.UFRGS.UFRGS_BI_SP_EN2	23.15%	36.80%	35.60%	27.40%
	4th	twente	10.2415/AH-TEL-BILI-X2EN-CLEF2008.TWENTE.FCW	22.78%	38.80%	35.20%	27.46%
	5th	jhu-apl	10.2415/AH-TEL-BILI-X2EN-CLEF2008.JHU-APL.JHUBIDEEN5	21.11%	35.20%	31.20%	22.26%
	Difference				61.77%	56.81%	60.89%
French	1st	chesire	10.2415/AH-TEL-BILI-X2FR-CLEF2008.CESHIRE.BKAHTELBDEFRTDT2FB	18.84%	30.00%	26.20%	17.53%
	2nd	chemnitz	10.2415/AH-TEL-BILI-X2FR-CLEF2008.CHEMNITZ.CUT_SIMPLE_EN2FR	17.54%	30.80%	26.00%	19.26%
	3rd	jhu-apl	10.2415/AH-TEL-BILI-X2FR-CLEF2008.JHU-APL.JHUBINLFR5	17.46%	28.00%	21.80%	14.33%
	4th	xerox	10.2415/AH-TEL-BILI-X2FR-CLEF2008.XEROX.GER_FRE_J	11.62%	17.20%	12.60%	8.73%
	5th	xerox-sas	10.2415/AH-TEL-BILI-X2FR-CLEF2008.XEROX-SAS.CACAOENGFREPLAIN	6.78%	11.20%	11.40%	8.13%
	Difference				177.87%	167.85%	129.82%
German	1st	jhu-apl	10.2415/AH-TEL-BILI-X2DE-CLEF2008.JHU-APL.JHUBIENDE5	18.98%	32.40%	26.60%	17.60%
	2nd	chemnitz	10.2415/AH-TEL-BILI-X2DE-CLEF2008.CHEMNITZ.CUT_MERGED_SIMPLE_EN2DE	18.51%	37.60%	32.60%	20.33%
	3rd	chesire	10.2415/AH-TEL-BILI-X2DE-CLEF2008.CESHIRE.BKAHTELBENDETDT2FB	15.56%	27.20%	23.40%	18.40%
	4th	xerox	10.2415/AH-TEL-BILI-X2DE-CLEF2008.XEROX.FRE_GER_J	12.05%	16.80%	16.20%	12.86%
	5th	karlsruhe	10.2415/AH-TEL-BILI-X2DE-CLEF2008.KARLSRUHE.AIFB_ONB_EN	6.67%	10.80%	9.80%	8.60%
	Difference				184.55%	248.14%	232.65%



TEL: Approaches to Bilingual IR

- Google online translation services
- Bilingual experiments performed almost as monolingual ones

Track	Rank	Participant	Experiment DOI	MAP	P@5	P@10	P@30
English	1st	chemnitz	10.2415/AH-TEL-BILI-X2EN-CLEF2008.CHEMNITZ.CUT_SIMPLE_DE2EN	34.15%	55.20%	50.20%	38.73%
	2nd	chesire	10.2415/AH-TEL-BILI-X2EN-CLEF2008.CESHIRE.BKAHTELBFRENTDT2FB	28.24%	46.40%	40.99%	30.73%
	3rd	ufrgs	10.2415/AH-TEL-BILI-X2EN-CLEF2008.UFRGS.UFRGS_BI_SP_EN2	23.15%	36.80%	35.60%	27.40%
	4th	twente	10.2415/AH-TEL-BILI-X2EN-CLEF2008.TWENTE.FCW	22.78%	38.80%	35.20%	27.46%
	5th	jhu-apl	10.2415/AH-TEL-BILI-X2EN-CLEF2008.JHU-APL.JHUBIDEEN5	21.11%	35.20%	31.20%	22.26%
	Difference				61.77%	56.81%	60.89%
French	1st	chesire	10.2415/AH-TEL-BILI-X2FR-CLEF2008.CESHIRE.BKAHTELBDEFRTDT2FB	18.84%	30.00%	26.20%	17.53%
	2nd	chemnitz	10.2415/AH-TEL-BILI-X2FR-CLEF2008.CHEMNITZ.CUT_SIMPLE_EN2FR	17.54%	30.80%	26.00%	19.26%
	3rd	jhu-apl	10.2415/AH-TEL-BILI-X2FR-CLEF2008.JHU-APL.JHUBINLFR5	17.46%	28.00%	21.80%	14.33%
	4th	xerox	10.2415/AH-TEL-BILI-X2FR-CLEF2008.XEROX.GER_FRE_J	11.62%	17.20%	12.60%	8.73%
	5th	xerox-sas	10.2415/AH-TEL-BILI-X2FR-CLEF2008.XEROX-SAS.CACAOENGFREPLAIN	6.78%	11.20%	11.40%	8.13%
	Difference				177.87%	167.85%	129.82%
German	1st	jhu-apl	10.2415/AH-TEL-BILI-X2DE-CLEF2008.JHU-APL.JHUBIENDE5	18.98%	32.40%	26.60%	17.60%
	2nd	chemnitz	10.2415/AH-TEL-BILI-X2DE-CLEF2008.CHEMNITZ.CUT_MERGED_SIMPLE_EN2DE	18.51%	37.60%	32.60%	20.33%
	3rd	chesire	10.2415/AH-TEL-BILI-X2DE-CLEF2008.CESHIRE.BKAHTELBENDETDT2FB	15.56%	27.20%	23.40%	18.40%
	4th	xerox	10.2415/AH-TEL-BILI-X2DE-CLEF2008.XEROX.FRE_GER_J	12.05%	16.80%	16.20%	12.86%
	5th	karlsruhe	10.2415/AH-TEL-BILI-X2DE-CLEF2008.KARLSRUHE.AIFB_ONB_EN	6.67%	10.80%	9.80%	8.60%
	Difference				184.55%	248.14%	232.65%



TEL: Approaches to Bilingual IR

LEC Power Translator MT system

Track	Rank	Participant	Experiment DOI	MAP	P@5	P@10	P@30
English	1st	chemnitz	10.2415/AH-TEL-BILI-X2EN-CLEF2008.CHEMNITZ.CUT_SIMPLE_DE2EN	34.15%	55.20%	50.20%	38.73%
	2nd	chesire	10.2415/AH-TEL-BILI-X2EN-CLEF2008.CESHIRE.BKAHTELBFRENTDT2FB	28.24%	46.40%	40.99%	30.73%
	3rd	ufrgs	10.2415/AH-TEL-BILI-X2EN-CLEF2008.UFRGS.UFRGS_BI_SP_EN2	23.15%	36.80%	35.60%	27.40%
	4th	twente	10.2415/AH-TEL-BILI-X2EN-CLEF2008.TWENTE.FCW	22.78%	38.80%	35.20%	27.46%
	5th	jhu-apl	10.2415/AH-TEL-BILI-X2EN-CLEF2008.JHU-APL.JHUBIDEEN5	21.11%	35.20%	31.20%	22.26%
	Difference				61.77%	56.81%	60.89%
French	1st	chesire	10.2415/AH-TEL-BILI-X2FR-CLEF2008.CESHIRE.BKAHTELBDEFRTDT2FB	18.84%	30.00%	26.20%	17.53%
	2nd	chemnitz	10.2415/AH-TEL-BILI-X2FR-CLEF2008.CHEMNITZ.CUT_SIMPLE_EN2FR	17.54%	30.80%	26.00%	19.26%
	3rd	jhu-apl	10.2415/AH-TEL-BILI-X2FR-CLEF2008.JHU-APL.JHUBINLFR5	17.46%	28.00%	21.80%	14.33%
	4th	xerox	10.2415/AH-TEL-BILI-X2FR-CLEF2008.XEROX.GER_FRE_J	11.62%	17.20%	12.60%	8.73%
	5th	xerox-sas	10.2415/AH-TEL-BILI-X2FR-CLEF2008.XEROX-SAS.CACAOENGFREPLAIN	6.78%	11.20%	11.40%	8.13%
	Difference				177.87%	167.85%	129.82%
German	1st	jhu-apl	10.2415/AH-TEL-BILI-X2DE-CLEF2008.JHU-APL.JHUBIENDE5	18.98%	32.40%	26.60%	17.60%
	2nd	chemnitz	10.2415/AH-TEL-BILI-X2DE-CLEF2008.CHEMNITZ.CUT_MERGED_SIMPLE_EN2DE	18.51%	37.60%	32.60%	20.33%
	3rd	chesire	10.2415/AH-TEL-BILI-X2DE-CLEF2008.CESHIRE.BKAHTELBENDEDT2FB	15.56%	27.20%	23.40%	18.40%
	4th	xerox	10.2415/AH-TEL-BILI-X2DE-CLEF2008.XEROX.FRE_GER_J	12.05%	16.80%	16.20%	12.86%
	5th	karlsruhe	10.2415/AH-TEL-BILI-X2DE-CLEF2008.KARLSRUHE.AIFB_ONB_EN	6.67%	10.80%	9.80%	8.60%
	Difference				184.55%	248.14%	232.65%



TEL: Approaches to Bilingual IR

- Traditional IR approach
- Multilinguality of the collection not considered
- Stop-words (Snowball)
- Snowball stemmers
- Probabilistic (Okapi/Zettair) Model modified to boost rare terms
- CLIR modeled as an association rule mining problem: cross-lingual equivalents have many co-occurrences in a parallel corpus (25% of the collection translated by Google translator)

Track	Rank	Participant	Experiment DOI	MAP	P@5	P@10	P@30
English	1st	chemnitz	10.2415/AH-TEL-BILI-X2EN-CLEF2008.CHEMNITZ.CUT_SIMPLE_DE2EN	34.15%	55.20%	50.20%	38.73%
	2nd	chesire	10.2415/AH-TEL-BILI-X2EN-CLEF2008.CESHIRE.BKAHTELBFRENTDT2FB	28.24%	46.40%	40.99%	30.73%
	3rd	ufrgs	10.2415/AH-TEL-BILI-X2EN-CLEF2008.UFRGS.UFRGS_BI_SP_EN2	23.15%	36.80%	35.60%	27.40%
	4th	twente	10.2415/AH-TEL-BILI-X2EN-CLEF2008.TWENTE.FCW	22.78%	38.80%	35.20%	27.46%
	5th	jhu-apl	10.2415/AH-TEL-BILI-X2EN-CLEF2008.JHU-APL.JHUBIDEEN5	21.11%	35.20%	31.20%	22.26%
	Difference				61.77%	56.81%	60.89%
French	1st	chesire	10.2415/AH-TEL-BILI-X2FR-CLEF2008.CESHIRE.BKAHTELBDEFRTDT2FB	18.84%	30.00%	26.20%	17.53%
	2nd	chemnitz	10.2415/AH-TEL-BILI-X2FR-CLEF2008.CHEMNITZ.CUT_SIMPLE_EN2FR	17.54%	30.80%	26.00%	19.26%
	3rd	jhu-apl	10.2415/AH-TEL-BILI-X2FR-CLEF2008.JHU-APL.JHUBINLFR5	17.46%	28.00%	21.80%	14.33%
	4th	xerox	10.2415/AH-TEL-BILI-X2FR-CLEF2008.XEROX.GER_FRE_J	11.62%	17.20%	12.60%	8.73%
	5th	xerox-sas	10.2415/AH-TEL-BILI-X2FR-CLEF2008.XEROX-SAS.CACAOENGFREPLAIN	6.78%	11.20%	11.40%	8.13%
	Difference				177.87%	167.85%	129.82%
German	1st	jhu-apl	10.2415/AH-TEL-BILI-X2DE-CLEF2008.JHU-APL.JHUBIENDE5	18.98%	32.40%	26.60%	17.60%
	2nd	chemnitz	10.2415/AH-TEL-BILI-X2DE-CLEF2008.CHEMNITZ.CUT_MERGED_SIMPLE_EN2DE	18.51%	37.60%	32.60%	20.33%
	3rd	chesire	10.2415/AH-TEL-BILI-X2DE-CLEF2008.CESHIRE.BKAHTELBENDETDT2FB	15.56%	27.20%	23.40%	18.40%
	4th	xerox	10.2415/AH-TEL-BILI-X2DE-CLEF2008.XEROX.FRE_GER_J	12.05%	16.80%	16.20%	12.86%
	5th	karlsruhe	10.2415/AH-TEL-BILI-X2DE-CLEF2008.KARLSRUHE.AIFB_ONB_EN	6.67%	10.80%	9.80%	8.60%
	Difference				184.55%	248.14%	232.65%



TEL: Approaches to Bilingual IR

- Traditional IR Approach
- Multilinguality of the collection not considered
- Stop-words (Snowball)
- Snowball stemmers and decompounding for Dutch queries
- Vector space (Lucene) Model
- WikiTranslate: map queries to Wikipedia concepts, follow the cross-lingual links translation of the concepts, use the terms in the translated concepts as query translation

Track	Rank	Participant	Experiment DOI	MAP	P@5	P@10	P@30
English	1st	chemnitz	10.2415/AH-TEL-BILI-X2EN-CLEF2008.CHEMNITZ.CUT_SIMPLE_DE2EN	34.15%	55.20%	50.20%	38.73%
	2nd	chesire	10.2415/AH-TEL-BILI-X2EN-CLEF2008.CESHIRE.BKAHTELBFRENTDT2FB	28.24%	46.40%	40.99%	30.73%
	3rd	ufrgs	10.2415/AH-TEL-BILI-X2EN-CLEF2008.UFRGS.UFRGS_BI_SP_EN2	23.15%	36.80%	35.60%	27.40%
	4th	twente	10.2415/AH-TEL-BILI-X2EN-CLEF2008.TWENTE.FCW	22.78%	38.80%	35.20%	27.46%
	5th	jhu-apl	10.2415/AH-TEL-BILI-X2EN-CLEF2008.JHU-APL.JHUBIDEEN5	21.11%	35.20%	31.20%	22.26%
	Difference				61.77%	56.81%	60.89%
French	1st	chesire	10.2415/AH-TEL-BILI-X2FR-CLEF2008.CESHIRE.BKAHTELBDEFRTDT2FB	18.84%	30.00%	26.20%	17.53%
	2nd	chemnitz	10.2415/AH-TEL-BILI-X2FR-CLEF2008.CHEMNITZ.CUT_SIMPLE_EN2FR	17.54%	30.80%	26.00%	19.26%
	3rd	jhu-apl	10.2415/AH-TEL-BILI-X2FR-CLEF2008.JHU-APL.JHUBINLFR5	17.46%	28.00%	21.80%	14.33%
	4th	xerox	10.2415/AH-TEL-BILI-X2FR-CLEF2008.XEROX.GER_FRE_J	11.62%	17.20%	12.60%	8.73%
	5th	xerox-sas	10.2415/AH-TEL-BILI-X2FR-CLEF2008.XEROX-SAS.CACAOENGFREPLAIN	6.78%	11.20%	11.40%	8.13%
	Difference				177.87%	167.85%	129.82%
German	1st	jhu-apl	10.2415/AH-TEL-BILI-X2DE-CLEF2008.JHU-APL.JHUBIENDE5	18.98%	32.40%	26.60%	17.60%
	2nd	chemnitz	10.2415/AH-TEL-BILI-X2DE-CLEF2008.CHEMNITZ.CUT_MERGED_SIMPLE_EN2DE	18.51%	37.60%	32.60%	20.33%
	3rd	chesire	10.2415/AH-TEL-BILI-X2DE-CLEF2008.CESHIRE.BKAHTELBENDETDT2FB	15.56%	27.20%	23.40%	18.40%
	4th	xerox	10.2415/AH-TEL-BILI-X2DE-CLEF2008.XEROX.FRE_GER_J	12.05%	16.80%	16.20%	12.86%
	5th	karlsruhe	10.2415/AH-TEL-BILI-X2DE-CLEF2008.KARLSRUHE.AIFB_ONB_EN	6.67%	10.80%	9.80%	8.60%
	Difference				184.55%	248.14%	232.65%



TEL: Approaches to Bilingual IR

- 5-grams of the source query were translated to the target language used a large aligned parallel corpus (the Official Journal of the European Commission)

Track	Rank	Participant	Experiment DOI	MAP	P@5	P@10	P@30
English	1st	chemnitz	10.2415/AH-TEL-BILI-X2EN-CLEF2008.CHEMNITZ.CUT_SIMPLE_DE2EN	34.15%	55.20%	50.20%	38.73%
	2nd	chesire	10.2415/AH-TEL-BILI-X2EN-CLEF2008.CESHIRE.BKAHTELBFRENTDT2FB	28.24%	46.40%	40.99%	30.73%
	3rd	ufrgs	10.2415/AH-TEL-BILI-X2EN-CLEF2008.UFRGS.UFRGS_BI_SP_EN2	23.15%	36.80%	35.60%	27.40%
	4th	twente	10.2415/AH-TEL-BILI-X2EN-CLEF2008.TWENTE.FCW	22.78%	38.80%	35.20%	27.46%
	5th	jhu-apl	10.2415/AH-TEL-BILI-X2EN-CLEF2008.JHU-APL.JHUBIDEEN5	21.11%	35.20%	31.20%	22.26%
	Difference				61.77%	56.81%	60.89%
French	1st	chesire	10.2415/AH-TEL-BILI-X2FR-CLEF2008.CESHIRE.BKAHTELBDEFRTDT2FB	18.84%	30.00%	26.20%	17.53%
	2nd	chemnitz	10.2415/AH-TEL-BILI-X2FR-CLEF2008.CHEMNITZ.CUT_SIMPLE_EN2FR	17.54%	30.80%	26.00%	19.26%
	3rd	jhu-apl	10.2415/AH-TEL-BILI-X2FR-CLEF2008.JHU-APL.JHUBINLFR5	17.46%	28.00%	21.80%	14.33%
	4th	xerox	10.2415/AH-TEL-BILI-X2FR-CLEF2008.XEROX.GER_FRE_J	11.62%	17.20%	12.60%	8.73%
	5th	xerox-sas	10.2415/AH-TEL-BILI-X2FR-CLEF2008.XEROX-SAS.CACAOENGFREPLAIN	6.78%	11.20%	11.40%	8.13%
	Difference				177.87%	167.85%	129.82%
German	1st	jhu-apl	10.2415/AH-TEL-BILI-X2DE-CLEF2008.JHU-APL.JHUBIENDE5	18.98%	32.40%	26.60%	17.60%
	2nd	chemnitz	10.2415/AH-TEL-BILI-X2DE-CLEF2008.CHEMNITZ.CUT_MERGED_SIMPLE_EN2DE	18.51%	37.60%	32.60%	20.33%
	3rd	chesire	10.2415/AH-TEL-BILI-X2DE-CLEF2008.CESHIRE.BKAHTELBENDETDT2FB	15.56%	27.20%	23.40%	18.40%
	4th	xerox	10.2415/AH-TEL-BILI-X2DE-CLEF2008.XEROX.FRE_GER_J	12.05%	16.80%	16.20%	12.86%
	5th	karlsruhe	10.2415/AH-TEL-BILI-X2DE-CLEF2008.KARLSRUHE.AIFB_ONB_EN	6.67%	10.80%	9.80%	8.60%
	Difference				184.55%	248.14%	232.65%



TEL: Approaches to Bilingual IR

- Multilingual probabilistic dictionary built using JRC-AC corpus plus approximate string matching techniques for lemmas not covered by JRC-AC combined with standard dictionary from ELRA

Track	Rank	Participant	Experiment DOI	MAP	P@5	P@10	P@30
English	1st	chemnitz	10.2415/AH-TEL-BILI-X2EN-CLEF2008.CHEMNITZ.CUT_SIMPLE_DE2EN	34.15%	55.20%	50.20%	38.73%
	2nd	chesire	10.2415/AH-TEL-BILI-X2EN-CLEF2008.CESHIRE.BKAHTELBFRENTDT2FB	28.24%	46.40%	40.99%	30.73%
	3rd	ufrgs	10.2415/AH-TEL-BILI-X2EN-CLEF2008.UFRGS.UFRGS_BI_SP_EN2	23.15%	36.80%	35.60%	27.40%
	4th	twente	10.2415/AH-TEL-BILI-X2EN-CLEF2008.TWENTE.FCW	22.78%	38.80%	35.20%	27.46%
	5th	jhu-apl	10.2415/AH-TEL-BILI-X2EN-CLEF2008.JHU-APL.JHUBIDEEN5	21.11%	35.20%	31.20%	22.26%
	Difference				61.77%	56.81%	60.89%
French	1st	chesire	10.2415/AH-TEL-BILI-X2FR-CLEF2008.CESHIRE.BKAHTELBDEFRTDT2FB	18.84%	30.00%	26.20%	17.53%
	2nd	chemnitz	10.2415/AH-TEL-BILI-X2FR-CLEF2008.CHEMNITZ.CUT_SIMPLE_EN2FR	17.54%	30.80%	26.00%	19.26%
	3rd	jhu-apl	10.2415/AH-TEL-BILI-X2FR-CLEF2008.JHU-APL.JHUBINLFR5	17.46%	28.00%	21.80%	14.33%
	4th	xerox	10.2415/AH-TEL-BILI-X2FR-CLEF2008.XEROX.GER_FRE_J	11.62%	17.20%	12.60%	8.73%
	5th	xerox-sas	10.2415/AH-TEL-BILI-X2FR-CLEF2008.XEROX-SAS.CACAOENGFREPLAIN	6.78%	11.20%	11.40%	8.13%
	Difference				177.87%	167.85%	129.82%
German	1st	jhu-apl	10.2415/AH-TEL-BILI-X2DE-CLEF2008.JHU-APL.JHUBIENDE5	18.98%	32.40%	26.60%	17.60%
	2nd	chemnitz	10.2415/AH-TEL-BILI-X2DE-CLEF2008.CHEMNITZ.CUT_MERGED_SIMPLE_EN2DE	18.51%	37.60%	32.60%	20.33%
	3rd	chesire	10.2415/AH-TEL-BILI-X2DE-CLEF2008.CESHIRE.BKAHTELBENDETDT2FB	15.56%	27.20%	23.40%	18.40%
	4th	xerox	10.2415/AH-TEL-BILI-X2DE-CLEF2008.XEROX.FRE_GER_J	12.05%	16.80%	16.20%	12.86%
	5th	karlsruhe	10.2415/AH-TEL-BILI-X2DE-CLEF2008.KARLSRUHE.AIFB_ONB_EN	6.67%	10.80%	9.80%	8.60%
	Difference				184.55%	248.14%	232.65%



TEL: Approaches to Bilingual IR

- Traditional IR Approach
- Multilinguality of the collection not considered
- Lemmatisation with XIP Xerox parsers and named entity recognition
- Vector Space (Lucene) Model
- Blind query expansion
- Translation by using internal resources and online dictionaries, such as Ergane, and disambiguation by using a word space model created with a random indexing approach

Track	Rank	Participant	Experiment DOI	MAP	P@5	P@10	P@30
English	1st	chemnitz	10.2415/AH-TEL-BILI-X2EN-CLEF2008.CHEMNITZ.CUT_SIMPLE_DE2EN	34.15%	55.20%	50.20%	38.73%
	2nd	chesire	10.2415/AH-TEL-BILI-X2EN-CLEF2008.CESHIRE.BKAHTELBFRENTDT2FB	28.24%	46.40%	40.99%	30.73%
	3rd	ufrgs	10.2415/AH-TEL-BILI-X2EN-CLEF2008.UFRGS.UFRGS_BI_SP_EN2	23.15%	36.80%	35.60%	27.40%
	4th	twente	10.2415/AH-TEL-BILI-X2EN-CLEF2008.TWENTE.FCW	22.78%	38.80%	35.20%	27.46%
	5th	jhu-apl	10.2415/AH-TEL-BILI-X2EN-CLEF2008.JHU-APL.JHUBIDEEN5	21.11%	35.20%	31.20%	22.26%
	Difference				61.77%	56.81%	60.89%
French	1st	chesire	10.2415/AH-TEL-BILI-X2FR-CLEF2008.CESHIRE.BKAHTELBDEFRTDT2FB	18.84%	30.00%	26.20%	17.53%
	2nd	chemnitz	10.2415/AH-TEL-BILI-X2FR-CLEF2008.CHEMNITZ.CUT_SIMPLE_EN2FR	17.54%	30.80%	26.00%	19.26%
	3rd	jhu-apl	10.2415/AH-TEL-BILI-X2FR-CLEF2008.JHU-APL.JHUBINLFR5	17.46%	28.00%	21.80%	14.33%
	4th	xerox	10.2415/AH-TEL-BILI-X2FR-CLEF2008.XEROX.GER_FRE_J	11.62%	17.20%	12.60%	8.73%
	5th	xerox-sas	10.2415/AH-TEL-BILI-X2FR-CLEF2008.XEROX-SAS.CACAOENGFREPLAIN	6.78%	11.20%	11.40%	8.13%
	Difference				177.87%	167.85%	129.82%
German	1st	jhu-apl	10.2415/AH-TEL-BILI-X2DE-CLEF2008.JHU-APL.JHUBIENDE5	18.98%	32.40%	26.60%	17.60%
	2nd	chemnitz	10.2415/AH-TEL-BILI-X2DE-CLEF2008.CHEMNITZ.CUT_MERGED_SIMPLE_EN2DE	18.51%	37.60%	32.60%	20.33%
	3rd	chesire	10.2415/AH-TEL-BILI-X2DE-CLEF2008.CESHIRE.BKAHTELBENDETD2FB	15.56%	27.20%	23.40%	18.40%
	4th	xerox	10.2415/AH-TEL-BILI-X2DE-CLEF2008.XEROX.FRE_GER_J	12.05%	16.80%	16.20%	12.86%
	5th	karlsruhe	10.2415/AH-TEL-BILI-X2DE-CLEF2008.KARLSRUHE.AIFB_ONB_EN	6.67%	10.80%	9.80%	8.60%
	Difference				184.55%	248.14%	232.65%



TEL: Approaches to Bilingual IR

- Traditional IR Approach
- Multilinguality of the collection considered: identification of the language of a record
- Stop-words
- Snowball stemmers
- Retrieval model based on Explicit Semantic Analysis (ESA, classification of a text wrt external categories)
- Cross-Lingual ESA: use Wikipedia concepts as external categories and exploit their cross-lingual links

Track	Rank	Participant	Experiment DOI	MAP	P@5	P@10	P@30
English	1st	chemnitz	10.2415/AH-TEL-BILI-X2EN-CLEF2008.CHEMNITZ.CUT_SIMPLE_DE2EN	34.15%	55.20%	50.20%	38.73%
	2nd	chesire	10.2415/AH-TEL-BILI-X2EN-CLEF2008.CESHIRE.BKAHTELBFRENTDT2FB	28.24%	46.40%	40.99%	30.73%
	3rd	ufrgs	10.2415/AH-TEL-BILI-X2EN-CLEF2008.UFRGS.UFRGS_BI_SP_EN2	23.15%	36.80%	35.60%	27.40%
	4th	twente	10.2415/AH-TEL-BILI-X2EN-CLEF2008.TWENTE.FCW	22.78%	38.80%	35.20%	27.46%
	5th	jhu-apl	10.2415/AH-TEL-BILI-X2EN-CLEF2008.JHU-APL.JHUBIDEEN5	21.11%	35.20%	31.20%	22.26%
	Difference				61.77%	56.81%	60.89%
French	1st	chesire	10.2415/AH-TEL-BILI-X2FR-CLEF2008.CESHIRE.BKAHTELBDEFRTDT2FB	18.84%	30.00%	26.20%	17.53%
	2nd	chemnitz	10.2415/AH-TEL-BILI-X2FR-CLEF2008.CHEMNITZ.CUT_SIMPLE_EN2FR	17.54%	30.80%	26.00%	19.26%
	3rd	jhu-apl	10.2415/AH-TEL-BILI-X2FR-CLEF2008.JHU-APL.JHUBINLFR5	17.46%	28.00%	21.80%	14.33%
	4th	xerox	10.2415/AH-TEL-BILI-X2FR-CLEF2008.XEROX.GER_FRE_J	11.62%	17.20%	12.60%	8.73%
	5th	xerox-sas	10.2415/AH-TEL-BILI-X2FR-CLEF2008.XEROX-SAS.CACAOENGFREPLAIN	6.78%	11.20%	11.40%	8.13%
	Difference				177.87%	167.85%	129.82%
German	1st	jhu-apl	10.2415/AH-TEL-BILI-X2DE-CLEF2008.JHU-APL.JHUBIENDE5	18.98%	32.40%	26.60%	17.60%
	2nd	chemnitz	10.2415/AH-TEL-BILI-X2DE-CLEF2008.CHEMNITZ.CUT_MERGED_SIMPLE_EN2DE	18.51%	37.60%	32.60%	20.33%
	3rd	chesire	10.2415/AH-TEL-BILI-X2DE-CLEF2008.CESHIRE.BKAHTELBENDETDT2FB	15.56%	27.20%	23.40%	18.40%
	4th	xerox	10.2415/AH-TEL-BILI-X2DE-CLEF2008.XEROX.FRE_GER_J	12.05%	16.80%	16.20%	12.86%
	5th	karlsruhe	10.2415/AH-TEL-BILI-X2DE-CLEF2008.KARLSRUHE.AIFB_ONB_EN	6.67%	10.80%	9.80%	8.60%
	Difference				184.55%	248.14%	232.65%



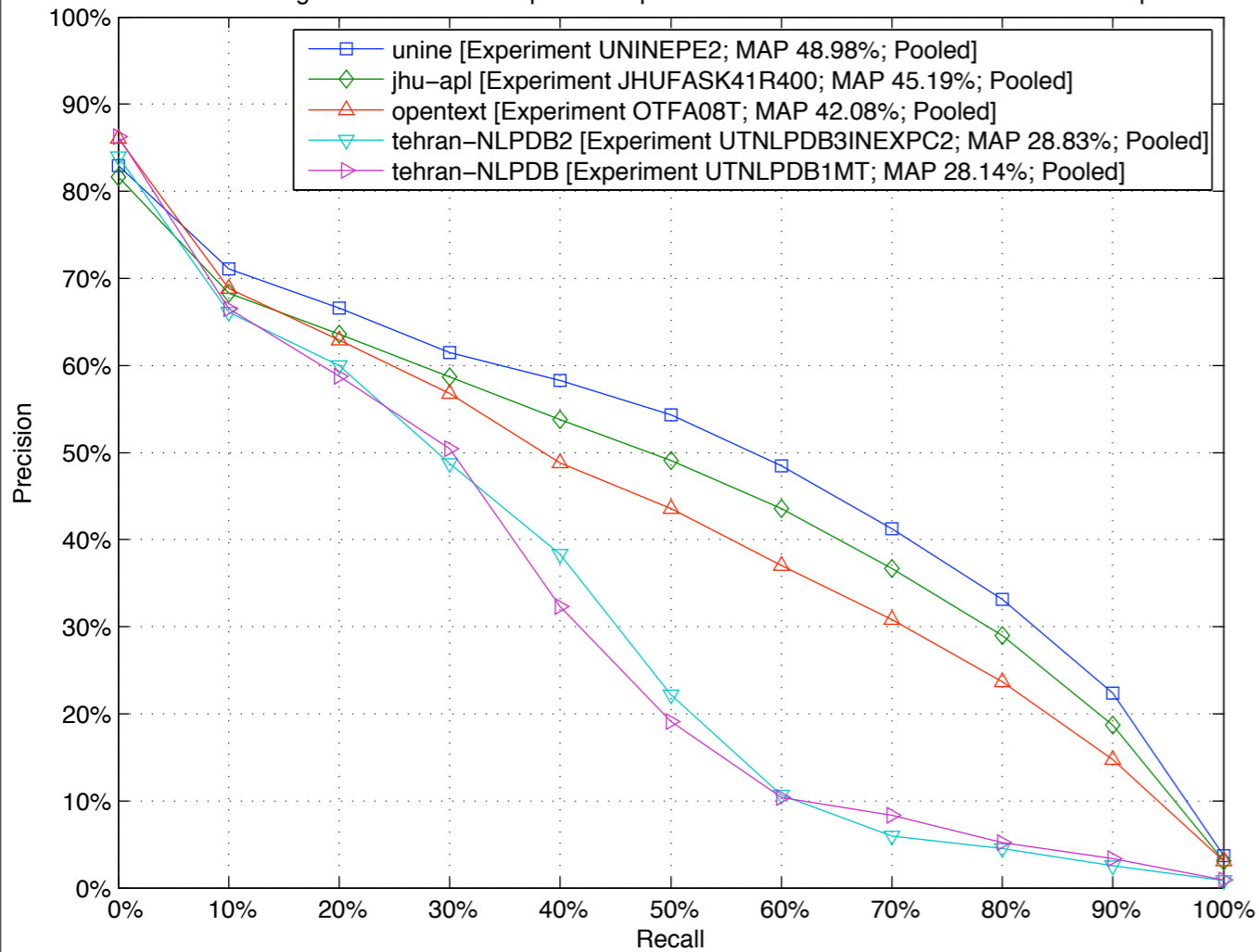
TEL: Approaches to Bilingual IR

Track	Rank	Participant	Experiment DOI	MAP	P@5	P@10	P@30
English	1st	chemnitz	10.2415/AH-TEL-BILI-X2EN-CLEF2008.CHEMNITZ.CUT_SIMPLE_DE2EN	34.15%	55.20%	50.20%	38.73%
	2nd	chesire	10.2415/AH-TEL-BILI-X2EN-CLEF2008.CESHIRE.BKAHTELBFRENTDT2FB	28.24%	46.40%	40.99%	30.73%
	3rd	ufrgs	10.2415/AH-TEL-BILI-X2EN-CLEF2008.UFRGS.UFRGS_BI_SP_EN2	23.15%	36.80%	35.60%	27.40%
	4th	twente	10.2415/AH-TEL-BILI-X2EN-CLEF2008.TWENTE.FCW	22.78%	38.80%	35.20%	27.46%
	5th	jhu-apl	10.2415/AH-TEL-BILI-X2EN-CLEF2008.JHU-APL.JHUBIDEEN5	21.11%	35.20%	31.20%	22.26%
	Difference				61.77%	56.81%	60.89%
French	1st	chesire	10.2415/AH-TEL-BILI-X2FR-CLEF2008.CESHIRE.BKAHTELBDEFRTDT2FB	18.84%	30.00%	26.20%	17.53%
	2nd	chemnitz	10.2415/AH-TEL-BILI-X2FR-CLEF2008.CHEMNITZ.CUT_SIMPLE_EN2FR	17.54%	30.80%	26.00%	19.26%
	3rd	jhu-apl	10.2415/AH-TEL-BILI-X2FR-CLEF2008.JHU-APL.JHUBINLFR5	17.46%	28.00%	21.80%	14.33%
	4th	xerox	10.2415/AH-TEL-BILI-X2FR-CLEF2008.XEROX.GER_FRE_J	11.62%	17.20%	12.60%	8.73%
	5th	xerox-sas	10.2415/AH-TEL-BILI-X2FR-CLEF2008.XEROX-SAS.CACAOENGFREPLAIN	6.78%	11.20%	11.40%	8.13%
	Difference				177.87%	167.85%	129.82%
German	1st	jhu-apl	10.2415/AH-TEL-BILI-X2DE-CLEF2008.JHU-APL.JHUBIENDE5	18.98%	32.40%	26.60%	17.60%
	2nd	chemnitz	10.2415/AH-TEL-BILI-X2DE-CLEF2008.CHEMNITZ.CUT_MERGED_SIMPLE_EN2DE	18.51%	37.60%	32.60%	20.33%
	3rd	chesire	10.2415/AH-TEL-BILI-X2DE-CLEF2008.CESHIRE.BKAHTELBENDEDT2FB	15.56%	27.20%	23.40%	18.40%
	4th	xerox	10.2415/AH-TEL-BILI-X2DE-CLEF2008.XEROX.FRE_GER_J	12.05%	16.80%	16.20%	12.86%
	5th	karlsruhe	10.2415/AH-TEL-BILI-X2DE-CLEF2008.KARLSRUHE.AIFB_ONB_EN	6.67%	10.80%	9.80%	8.60%
	Difference				184.55%	248.14%	232.65%

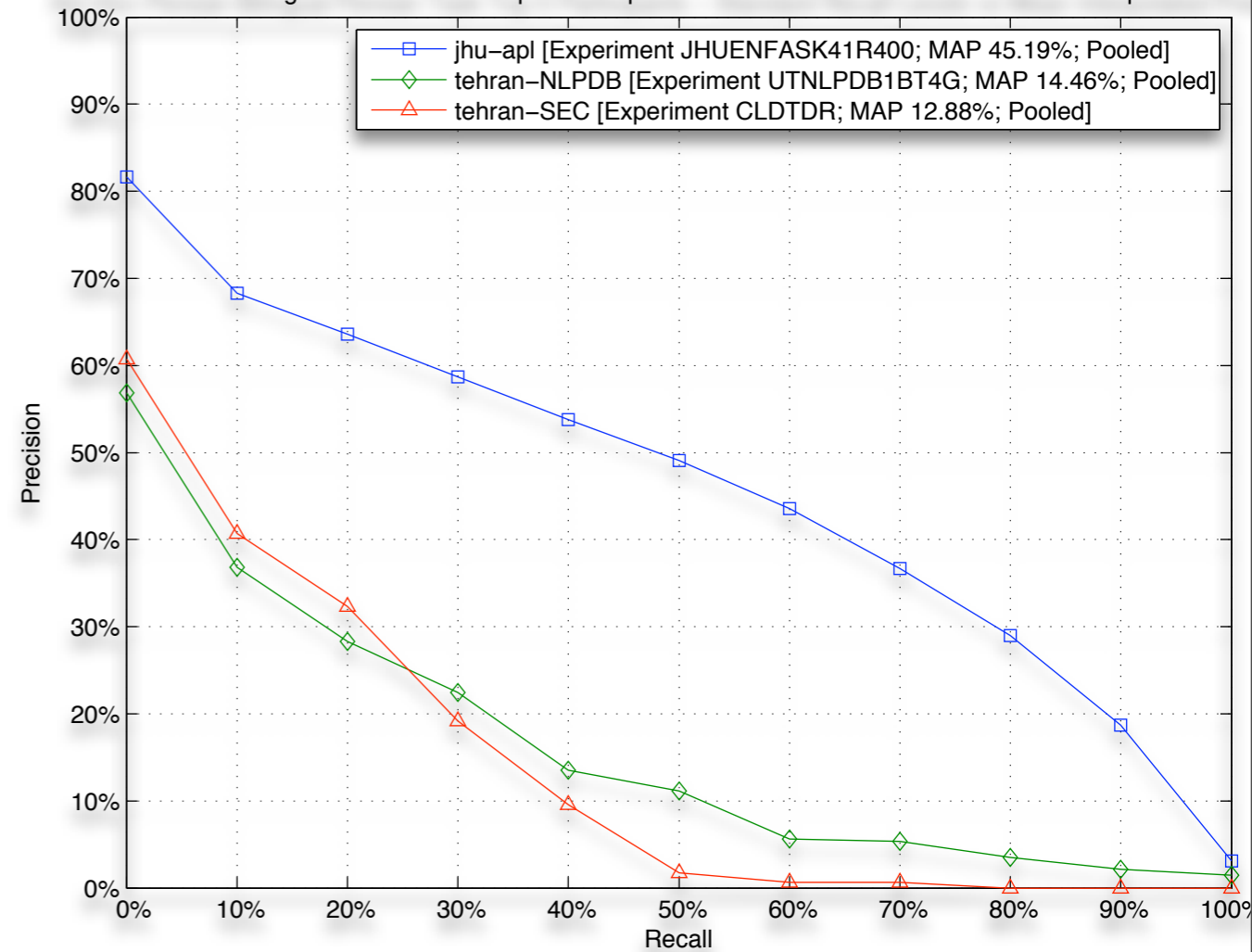


Persian

Ad-Hoc Persian Monolingual Persian Task Top 5 Participants – Standard Recall Levels vs Mean Interpolated Precision



Ad-Hoc Persian Bilingual Persian Task Top 5 Participants – Standard Recall Levels vs Mean Interpolated Precision

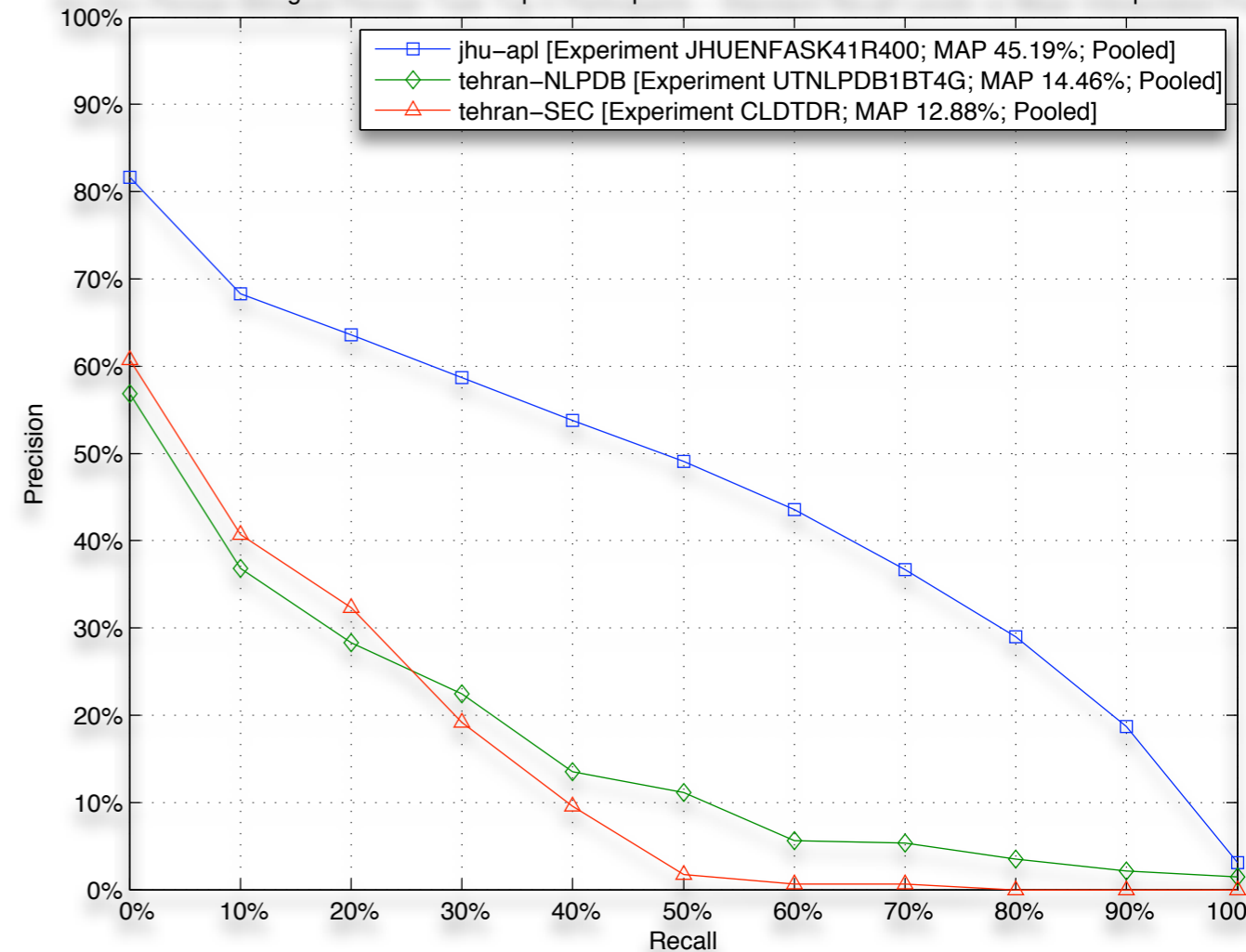
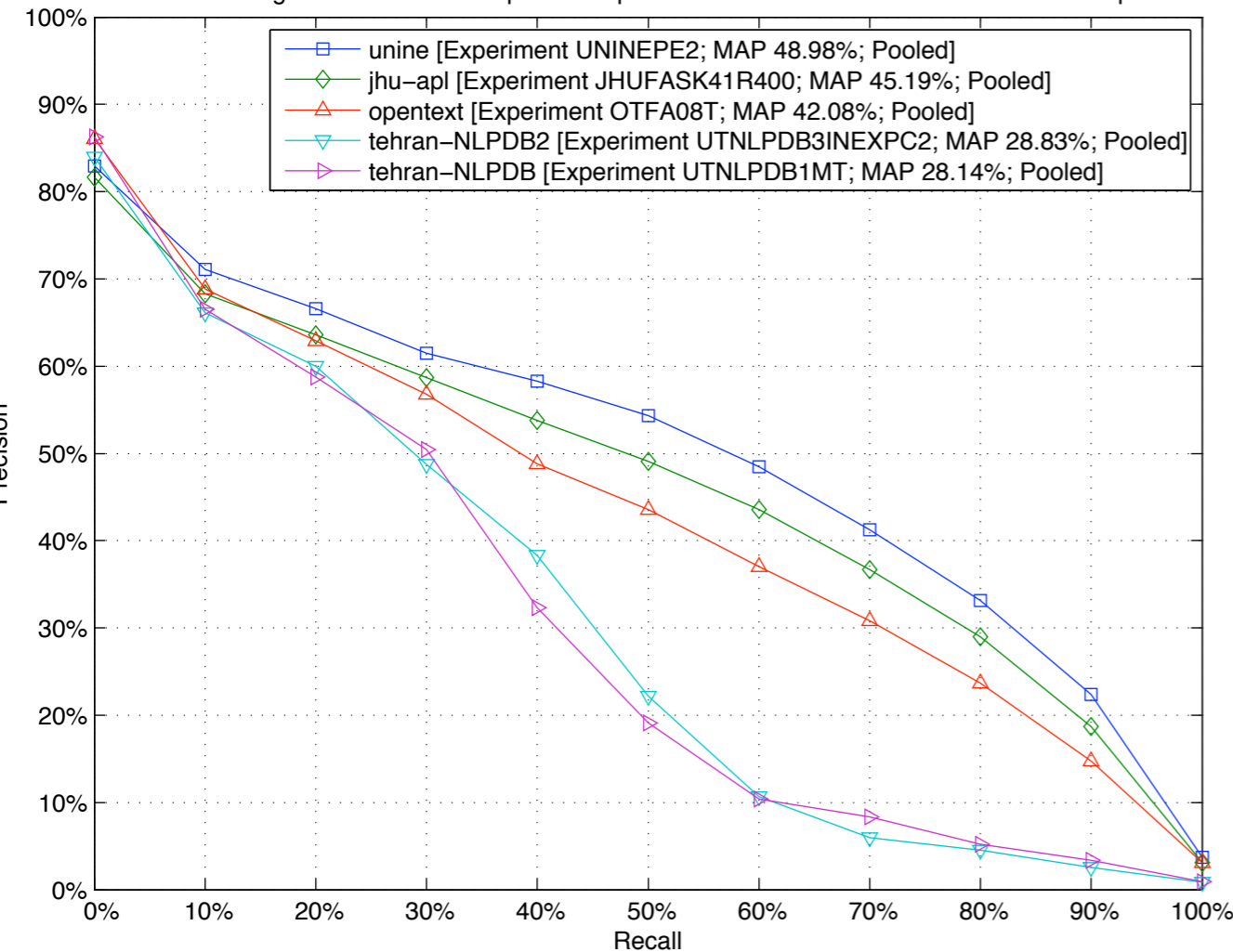




Persian

**Bilingual is
92%
of monolingual**

Ad-Hoc Persian Monolingual Persian Task Top 5 Participants – Standard Recall Levels vs Mean Interpolated Precision | Ad-Hoc Persian Bilingual Persian Task Top 5 Participants – Standard Recall Levels vs Mean Interpolated Precision





Persian: Approaches to Monolingual IR

Track	Rank	Participant	Experiment DOI	MAP	P@5	P@10	P@30
Monolingual	1st	unine	10.2415/AH-PERSIAN-MONO-FA-CLEF2008.UNINE.UNINEPE2	48.98%	69.20%	66.99%	59.53%
	2nd	jhu-apl	10.2415/AH-PERSIAN-MONO-FA-CLEF2008.JHU-APL.JHUFASK41R400	45.19%	65.20%	60.20%	56.20%
	3rd	opentext	10.2415/AH-PERSIAN-MONO-FA-CLEF2008.OPENTEXT.OTFA08T	42.08%	69.40%	62.59%	56.26%
	4th	tehran-nlpdb2	10.2415/AH-PERSIAN-MONO-FA-CLEF2008.TEHRAN-NLPDB2.UTNLPDB3INEXPC2	28.83%	65.60%	62.00%	53.53%
	5th	tehran-nlpdb	10.2415/AH-PERSIAN-MONO-FA-CLEF2008.TEHRAN-NLPDB.UTNLPDB1MT	28.14%	62.80%	61.80%	52.93%
	Difference				74.05%	10.20%	11.27%



Persian: Approaches to Monolingual IR

- Stop-words (884 terms)
- Light stemmer
- Indexing unit: 4-grams or words
- Probabilistic (Okapi and DFR) and Language Models
- Data fusion with Z-score
- Blind query expansion (Rocchio and idf, 5-10 documents, 10-50 terms)

Track	Rank	Participant	Experiment DOI	MAP	P@5	P@10	P@30
Monolingual	1st	unine	10.2415/AH-PERSIAN-MONO-FA-CLEF2008.UNINE.UNINEPE2	48.98%	69.20%	66.99%	59.53%
	2nd	jhu-apl	10.2415/AH-PERSIAN-MONO-FA-CLEF2008.JHU-APL.JHUFASK41R400	45.19%	65.20%	60.20%	56.20%
	3rd	opentext	10.2415/AH-PERSIAN-MONO-FA-CLEF2008.OPENTEXT.OTFA08T	42.08%	69.40%	62.59%	56.26%
	4th	tehran-nlpdb2	10.2415/AH-PERSIAN-MONO-FA-CLEF2008.TEHRAN-NLPDB2.UTNLPDB3INEXPC2	28.83%	65.60%	62.00%	53.53%
	5th	tehran-nlpdb	10.2415/AH-PERSIAN-MONO-FA-CLEF2008.TEHRAN-NLPDB.UTNLPDB1MT	28.14%	62.80%	61.80%	52.93%
	Difference				74.05%	10.20%	11.27%



Persian: Approaches to Monolingual IR

- 5-grams, skip-grams stemmers, morphological stemmer, no stemmer
- Language model with smoothing between term frequencies and corpus frequencies
- Blind query expansion (skip-grams required more)

Track	Rank	Participant	Experiment DOI	MAP	P@5	P@10	P@30
Monolingual	1st	unine	10.2415/AH-PERSIAN-MONO-FA-CLEF2008.UNINE.UNINEPE2	48.98%	69.20%	66.99%	59.53%
	2nd	jhu-apl	10.2415/AH-PERSIAN-MONO-FA-CLEF2008.JHU-APL.JHUFASK41R400	45.19%	65.20%	60.20%	56.20%
	3rd	opentext	10.2415/AH-PERSIAN-MONO-FA-CLEF2008.OPENTEXT.OTFA08T	42.08%	69.40%	62.59%	56.26%
	4th	tehran-nlpdb2	10.2415/AH-PERSIAN-MONO-FA-CLEF2008.TEHRAN-NLPDB2.UTNLPDB3INEXPC2	28.83%	65.60%	62.00%	53.53%
	5th	tehran-nlpdb	10.2415/AH-PERSIAN-MONO-FA-CLEF2008.TEHRAN-NLPDB.UTNLPDB1MT	28.14%	62.80%	61.80%	52.93%
	Difference				74.05%	10.20%	11.27%



Persian: Approaches to Monolingual IR

- linguistic, algorithmic, and 4-grams stemmers
- Language model with smoothing between term frequencies and corpus frequencies
- Blind query expansion (not top 5 ranked)

Track	Rank	Participant	Experiment DOI	MAP	P@5	P@10	P@30
Monolingual	1st	unine	10.2415/AH-PERSIAN-MONO-FA-CLEF2008.UNINE.UNINEPE2	48.98%	69.20%	66.99%	59.53%
	2nd	jhu-apl	10.2415/AH-PERSIAN-MONO-FA-CLEF2008.JHU-APL.JHUFASK41R400	45.19%	65.20%	60.20%	56.20%
	3rd	opentext	10.2415/AH-PERSIAN-MONO-FA-CLEF2008.OPENTEXT.OTFA08T	42.08%	69.40%	62.59%	56.26%
	4th	tehran-nlpdb2	10.2415/AH-PERSIAN-MONO-FA-CLEF2008.TEHRAN-NLPDB2.UTNLPDB3INEXPC2	28.83%	65.60%	62.00%	53.53%
	5th	tehran-nlpdb	10.2415/AH-PERSIAN-MONO-FA-CLEF2008.TEHRAN-NLPDB.UTNLPDB1MT	28.14%	62.80%	61.80%	52.93%
	Difference				74.05%	10.20%	11.27%



Persian: Approaches to Monolingual IR

- 9 Different retrieval models: probabilistic (Okapi, Okapi+DFR), DFR, tf/idf
- Data fusion with different retrieval models with a normal distribution-base weighting operator
- No significative improvement over the single models

Track	Rank	Participant	Experiment DOI	MAP	P@5	P@10	P@30
Monolingual	1st	unine	10.2415/AH-PERSIAN-MONO-FA-CLEF2008.UNINE.UNINEPE2	48.98%	69.20%	66.99%	59.53%
	2nd	jhu-apl	10.2415/AH-PERSIAN-MONO-FA-CLEF2008.JHU-APL.JHUFASK41R400	45.19%	65.20%	60.20%	56.20%
	3rd	opentext	10.2415/AH-PERSIAN-MONO-FA-CLEF2008.OPENTEXT.OTFA08T	42.08%	69.40%	62.59%	56.26%
	4th	tehran-nlpdb2	10.2415/AH-PERSIAN-MONO-FA-CLEF2008.TEHRAN-NLPDB2.UTNLPDB3INEXPC2	28.83%	65.60%	62.00%	53.53%
	5th	tehran-nlpdb	10.2415/AH-PERSIAN-MONO-FA-CLEF2008.TEHRAN-NLPDB.UTNLPDB1MT	28.14%	62.80%	61.80%	52.93%
	Difference				74.05%	10.20%	11.27%



Persian: Approaches to Monolingual IR

- Stop-words (796 terms) manually created
- Light stemmer, 3/4/5-grams
- Vector space (Lucene) Model and Language (Lemur) Model

Track	Rank	Participant	Experiment DOI	MAP	P@5	P@10	P@30
Monolingual	1st	unine	10.2415/AH-PERSIAN-MONO-FA-CLEF2008.UNINE.UNINEPE2	48.98%	69.20%	66.99%	59.53%
	2nd	jhu-apl	10.2415/AH-PERSIAN-MONO-FA-CLEF2008.JHU-APL.JHUFASK41R400	45.19%	65.20%	60.20%	56.20%
	3rd	opentext	10.2415/AH-PERSIAN-MONO-FA-CLEF2008.OPENTEXT.OTFA08T	42.08%	69.40%	62.59%	56.26%
	4th	tehran-nlpdb2	10.2415/AH-PERSIAN-MONO-FA-CLEF2008.TEHRAN-NLPDB2.UTNLPDB3INEXPC2	28.83%	65.60%	62.00%	53.53%
	5th	tehran-nlpdb	10.2415/AH-PERSIAN-MONO-FA-CLEF2008.TEHRAN-NLPDB.UTNLPDB1MT	28.14%	62.80%	61.80%	52.93%
	Difference				74.05%	10.20%	11.27%



Persian: Approaches to Monolingual IR

Track	Rank	Participant	Experiment DOI	MAP	P@5	P@10	P@30
Monolingual	1st	unine	10.2415/AH-PERSIAN-MONO-FA-CLEF2008.UNINE.UNINEPE2	48.98%	69.20%	66.99%	59.53%
	2nd	jhu-apl	10.2415/AH-PERSIAN-MONO-FA-CLEF2008.JHU-APL.JHUFASK41R400	45.19%	65.20%	60.20%	56.20%
	3rd	opentext	10.2415/AH-PERSIAN-MONO-FA-CLEF2008.OPENTEXT.OTFA08T	42.08%	69.40%	62.59%	56.26%
	4th	tehran-nlpdb2	10.2415/AH-PERSIAN-MONO-FA-CLEF2008.TEHRAN-NLPDB2.UTNLPDB3INEXPC2	28.83%	65.60%	62.00%	53.53%
	5th	tehran-nlpdb	10.2415/AH-PERSIAN-MONO-FA-CLEF2008.TEHRAN-NLPDB.UTNLPDB1MT	28.14%	62.80%	61.80%	52.93%
	Difference				74.05%	10.20%	11.27%



Persian: Approaches to Bilingual IR

Track	Rank	Participant	Experiment DOI	MAP	P@5	P@10	P@30
Bilingual	1st	jhu-apl	10.2415/AH-PERSIAN-BILI-X2FA-CLEF2008.JHU-APL.JHUENFASK41R400	45.19%	65.20%	60.59%	56.20%
	2nd	tehran-nlpdb	10.2415/AH-PERSIAN-BILI-X2FA-CLEF2008.TEHRAN-NLPDB.UTNLPDB1BT4G	14.45%	37.60%	33.60%	30.20%
	3rd	tehran-sec	10.2415/AH-PERSIAN-BILI-X2FA-CLEF2008.TEHRAN-SEC.CLDTDR	12.88%	40.00%	40.40%	33.80%
	4th	—	—	—	—	—	—
	5th	—	—	—	—	—	—
	Difference				250.85%	73.40%	80.32%



Persian: Approaches to Bilingual IR

● Online MT system (Parstranslator)

Track	Rank	Participant	Experiment DOI	MAP	P@5	P@10	P@30
Bilingual	1st	jhu-apl	10.2415/AH-PERSIAN-BILI-X2FA-CLEF2008.JHU-APL.JHUENFASK41R400	45.19%	65.20%	60.59%	56.20%
	2nd	tehran-nlpdb	10.2415/AH-PERSIAN-BILI-X2FA-CLEF2008.TEHRAN-NLPDB.UTNLPDB1BT4G	14.45%	37.60%	33.60%	30.20%
	3rd	tehran-sec	10.2415/AH-PERSIAN-BILI-X2FA-CLEF2008.TEHRAN-SEC.CLDTDR	12.88%	40.00%	40.40%	33.80%
	4th	—	—	—	—	—	—
	5th	—	—	—	—	—	—
	Difference				250.85%	73.40%	80.32%



Persian: Approaches to Bilingual IR

- Language (Lemur) Model
- Bilingual Persian-English Dictionary (more than 50,000 terms)
- Probabilistic query translation based on the co-occurrence of terms in the target collection

Track	Rank	Participant	Experiment DOI	MAP	P@5	P@10	P@30
Bilingual	1st	jhu-apl	10.2415/AH-PERSIAN-BILI-X2FA-CLEF2008.JHU-APL.JHUENFASK41R400	45.19%	65.20%	60.59%	56.20%
	2nd	tehran-nlpdb	10.2415/AH-PERSIAN-BILI-X2FA-CLEF2008.TEHRAN-NLPDB.UTNLPDB1BT4G	14.45%	37.60%	33.60%	30.20%
	3rd	tehran-sec	10.2415/AH-PERSIAN-BILI-X2FA-CLEF2008.TEHRAN-SEC.CLDTDR	12.88%	40.00%	40.40%	33.80%
	4th	—	—	—	—	—	—
	5th	—	—	—	—	—	—
	Difference				250.85%	73.40%	80.32%



Persian: Approaches to Bilingual IR

- Vector Space (Lucene) Model
- Document translation by using MT system (Shiraz, open source)

Track	Rank	Participant	Experiment DOI	MAP	P@5	P@10	P@30
Bilingual	1st	jhu-apl	10.2415/AH-PERSIAN-BILI-X2FA-CLEF2008.JHU-APL.JHUENFASK41R400	45.19%	65.20%	60.59%	56.20%
	2nd	tehran-nlpdb	10.2415/AH-PERSIAN-BILI-X2FA-CLEF2008.TEHRAN-NLPDB.UTNLPDB1BT4G	14.45%	37.60%	33.60%	30.20%
	3rd	tehran-sec	10.2415/AH-PERSIAN-BILI-X2FA-CLEF2008.TEHRAN-SEC.CLDTDR	12.88%	40.00%	40.40%	33.80%
	4th	—	—	—	—	—	—
	5th	—	—	—	—	—	—
	Difference				250.85%	73.40%	80.32%



Persian: Approaches to Bilingual IR

Track	Rank	Participant	Experiment DOI	MAP	P@5	P@10	P@30
Bilingual	1st	jhu-apl	10.2415/AH-PERSIAN-BILI-X2FA-CLEF2008.JHU-APL.JHUENFASK41R400	45.19%	65.20%	60.59%	56.20%
	2nd	tehran-nlpdb	10.2415/AH-PERSIAN-BILI-X2FA-CLEF2008.TEHRAN-NLPDB.UTNLPDB1BT4G	14.45%	37.60%	33.60%	30.20%
	3rd	tehran-sec	10.2415/AH-PERSIAN-BILI-X2FA-CLEF2008.TEHRAN-SEC.CLDTDR	12.88%	40.00%	40.40%	33.80%
	4th	—	—	—	—	—	—
	5th	—	—	—	—	—	—
	Difference				250.85%	73.40%	80.32%



Conclusions

- ❑ Encouraging participation to both tasks and interesting results have been achieved

- ❑ The experience gained this year will be very useful to further tune the tasks

- ❑ TEL Task:
 - traditional IR approaches seem to work well and achieve good results
 - only two groups have exploited the inherent multilinguality of the data
 - almost no group has exploited the semi-structured nature of the data or used the subject headings



Conclusions

- ❑ Encouraging participation to both tasks and interesting results have been achieved

- ❑ The experience gained this year will be very useful to further tune the tasks

- ❑ TEL Task:
 - traditional IR approaches seem to work well and achieve good results
 - only two groups have exploited the inherent multilinguality of the data
 - almost no group has exploited the semi-structured nature of the data or used the subject headings

We need to do more



Next: Grid Experiments on Friday at Break-out

