

# The BOLT IR Test Collections of Multilingual Passage Retrieval from Discussion Forums

Ian Soboroff, National Institute of Standards and Technology, ian.soboroff@nist.gov

Kira Griffitt and Stephanie Strassel, Linguistic Data Consortium, {kiragrif, strassel}@ldc.upenn.edu

LDC release is in  
review, planned for  
late 2016 or early  
2017 availability.

## TASK:

Systems retrieve short passages and translate them to English.

The user can drill down to the original post, so its ok if the passages don't have all the context needed.

**What are the consequences of increases in gasoline prices?**

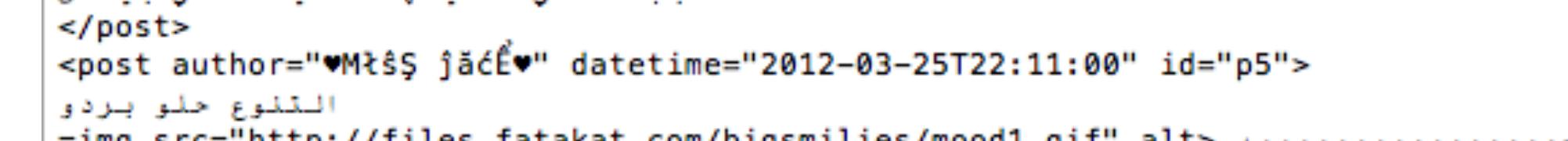
**bolt-arz-DF-252-200146-11454652**  
Pumping 21 million liters a day and no increase in prices follow:  
Farouk Abdul Aziz - Harbi perm - Safa al-Zayyat - Mohammed Jasmine - Shaalan Abdel Sadiq Mamdouh Ramadan - mi Mujahid - Ahmed al-Saadi

**bolt-arz-DF-207-186121-473598**  
The last of which was raising gasoline prices by 30% and the claim that raising the price of a liter of gasoline will not lead to an increase in the prices of commodities and other services, and it is a lie one can not swallow it, because the price

**bolt-arz-DF-169-181081-58202**  
long as the oil price rose in this way, it is not reasonable, keeping the price of a ton of diesel at one 600th pounds, as expected to increase petrol prices "90 and 92 and 95," and perhaps the price of gasoline 80 remains unchanged, with increasing

**bolt-arz-DF-175-182187-16036562**  
these products and not resort to storing petroleum products, whether diesel or gasoline, as this increases the aggravation of the crisis and not a solution, stressing that throughout the hours of the night didn't stop cars from pumping gasoline and

**bolt-arz-DF-198-195929-7932142**  
I suggest that the price of a liter of gasoline from all types of price

```
<?xml version="1.0" encoding="UTF-8"?>
<doc id="bolt-arz-DF-175-182187-16036560">
<headline>
باريت الاخوان ينهمنا ايه اللي يحصل
</headline>
<post author="14" datetime="2012-03-25T22:02:00" id="p1">
السلام عليكم ورحمة الله وبركاته
ازيكم بآفاقوكات
</post>
<post author="am_hassan_w_ghad" datetime="2012-03-25T22:08:00" id="p2">
يايايني اخلي كل حاجة
</post>
<post author="marmar92" datetime="2012-03-25T22:08:00" id="p3">
انا واحدة من الناس اللي ضد الاخوان وقد تفتقيرهم
</post>
<post author="14" datetime="2012-03-25T22:11:00" id="p4">
لابد لنفس حد ينهملا ايه اللي يحصل
</post>
<post author="M25_jacf" datetime="2012-03-25T22:11:00" id="p5">
الشلون جلو بردو

</post>
<post author="ha_ahmed" datetime="2012-03-25T22:16:00" id="p7">
انا فلدت الله فيهم خلاص
</post>
<post author="no_one_like_rere" datetime="2012-03-25T22:17:00" id="p8">
مش يعجم ابدا مع ان جوزي منهم
</post>
<post author="صاحب_اللذوق_الرقيق" datetime="2012-03-25T22:21:00" id="p9">
السلام عليكم ورحمة الله وبركاته
</post>
```

## Complex queries:

“asks-about”:

- person
- location
- organization
- event
- movement
- abstract entity

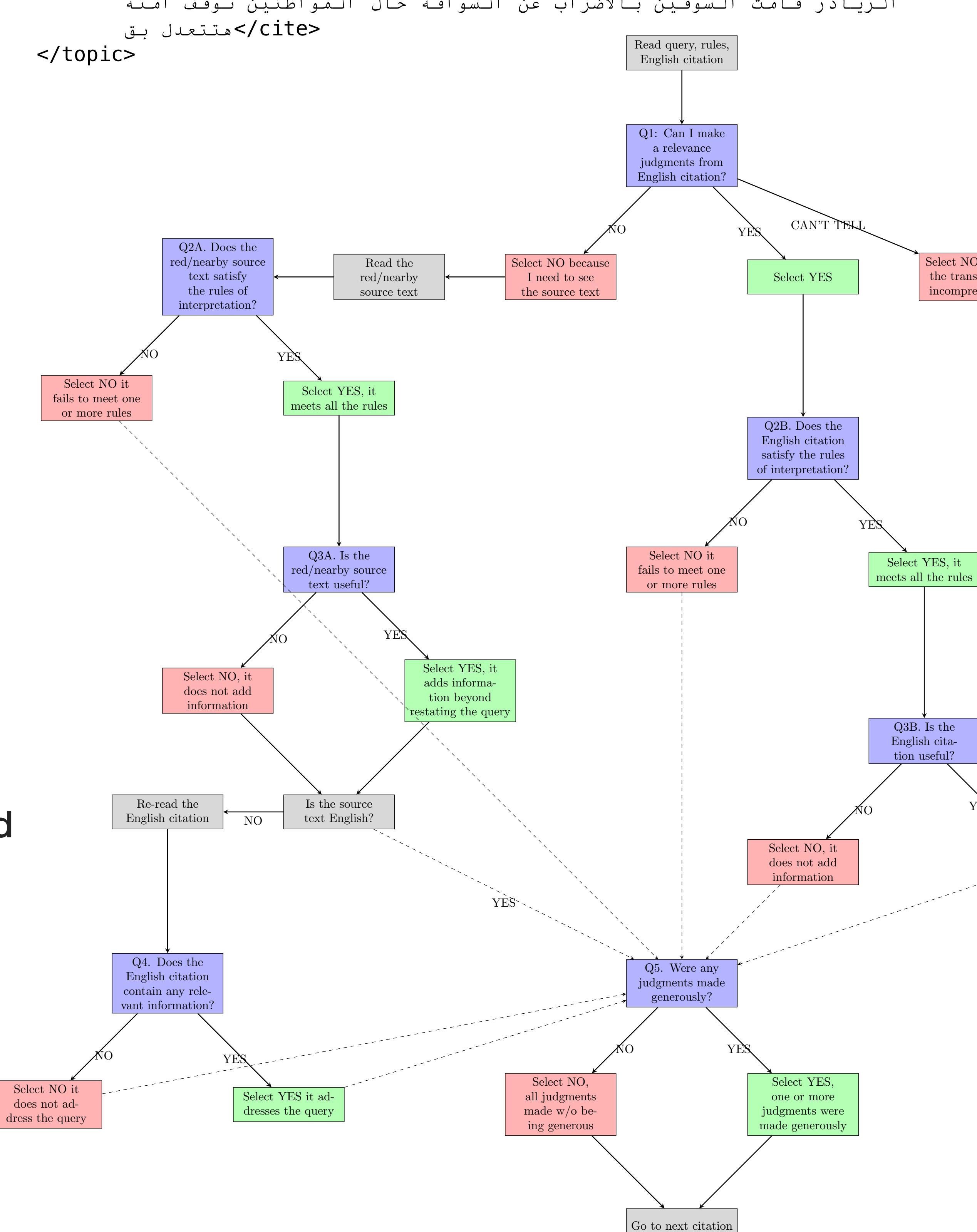
“asks-for”:

- statements or opinions
- relationships among ...
- effects of ...
- information about ...
- members of ...

Retrieved passages were pooled from BOLT evaluation systems and anonymized for assessment.

Native speakers at LDC assessed the pooled passages for relevance and translation usability.

```
<topic number="BIR_300072">
<query>What are the consequences of increases in gasoline prices?</query>
<description>This query asks for the effects of increases in gasoline prices.</description>
<language-target lang="arz"/>
<properties>
<asks-about target="product"/>
<asks-for response="effects of"/>
<languages eng="F" arz="T" cmn="F"/>
</properties>
<rule number="1">Answers must be about the increase in gasoline prices, not other commodities.</rule>
<rule number="2">Answers must be about the consequences of the increase in gasoline prices, not the causes.</rule>
<cite number="1" thread="bolt-arz-DF-204-185979-1495313" post="" offset="" length="59" rel="yes" dialect="no-ARZ">ارتفاع أسعار البنزين وانتباكات في محطات البنزين</cite>
<cite number="2" thread="bolt-arz-DF-274-201750-14002443" post="1" offset="1" length="133" rel="yes" dialect="no-ARZ">نتيجة لارتفاع أسعار البنزين فقد قام ملاك محطات البنزين بإغلاق محطاتهم وإطفاء الأنوار</cite>
<cite number="3" thread="bolt-arz-DF-217-194296-6679542" post="1" offset="71" length="220" rel="yes" dialect="ARZ">رتفعت أسعار البنزين وقلت نسبة وارتفاع تمن الاجرة وأضرر السواقين قولنا الثورة قامت وكل حاجة منتعد لكن للاسف دلوكتنا ثبات شفاعة لوكات اتعالنت قام مجال الشرطة بضبط متعدد لكن مصيفيت بنزين هورية وبيعها في السوق السوداء والقارة فيه</cite>
<cite number="4" thread="bolt-arz-DF-217-194296-6679542" post="1" offset="293" length="160" rel="yes" dialect="ARZ">كان عندما ارتفع سعر البنزين وقل قات السوقين برفع الاجة وعندما امتنع المواطنون عن دفع الزبادار قات السوقين بالاضراب عن السواقة حال المواطنون توقف امته</cite>
</topic>
```

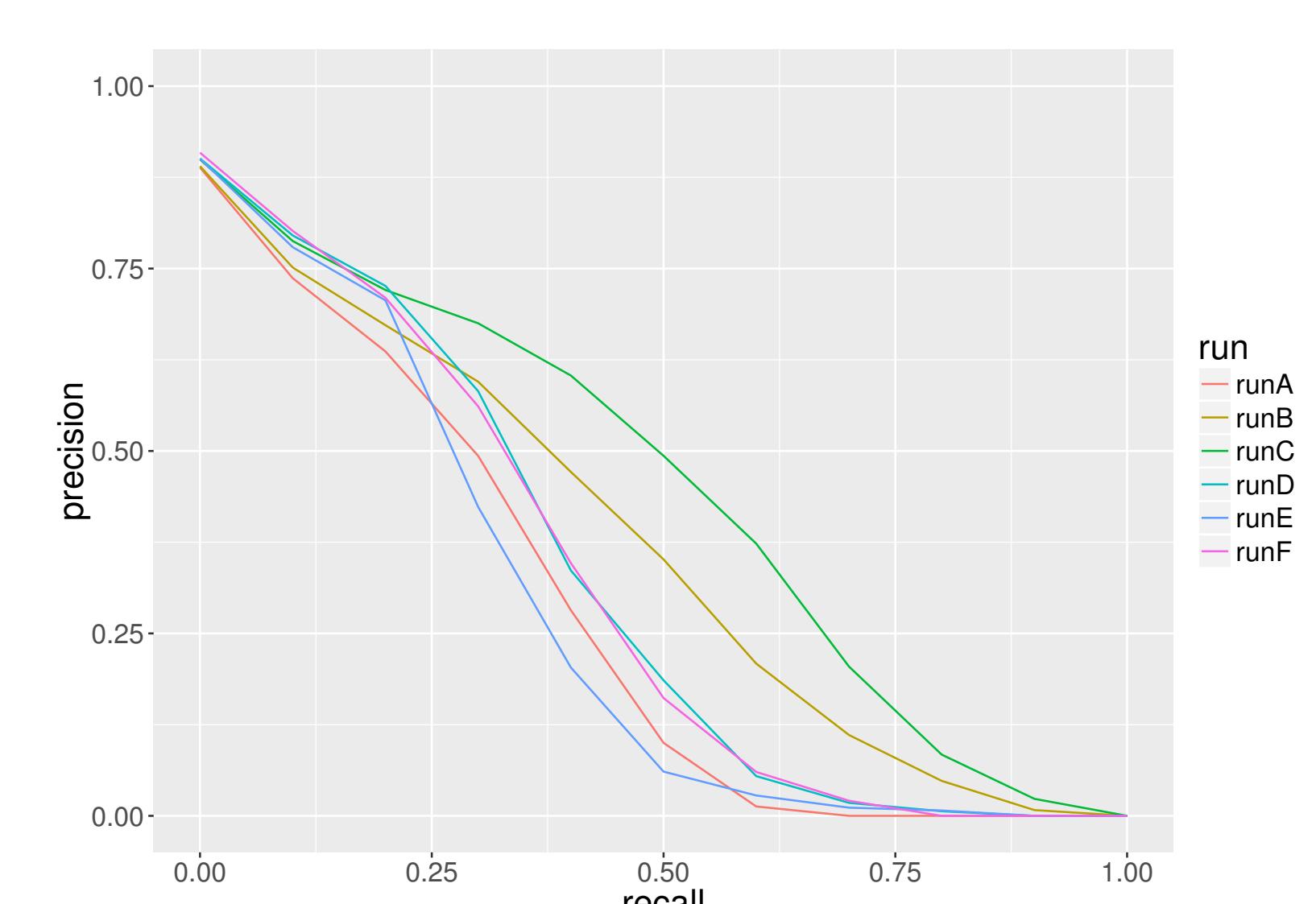
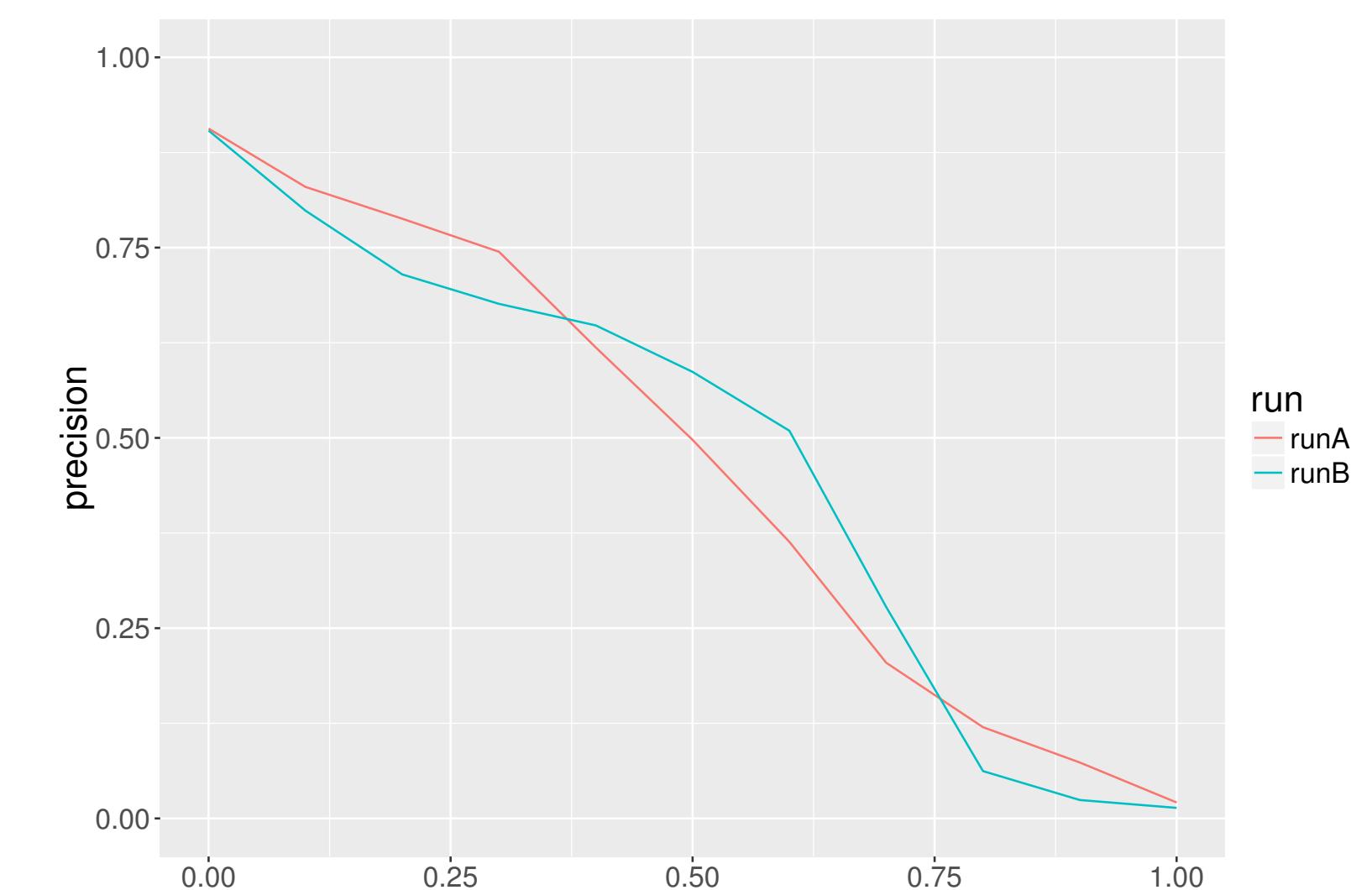


## Vital statistics

Language	English	Chinese	Arabic
tokens	694,914,443	1,050,254,101	616,719,471
threads	457,907	789,077	773,861
average msgs/thread	13.23	50.77	30.82
average thread size in bytes	14,651	5,769	11,380

Topic set	P2	P3
number of topics	100	150
average relevant passages per topic	114.56	252.46

## Baseline scores



This material is based upon work supported by the Defense Advanced Research Projects Agency (DARPA) under Contract No. HR0011-11-C-0145. The content does not necessarily reflect the position or the policy of the Government, and no official endorsement should be inferred.