رومینا اعتضادی - 97443301

ابزار برای Normalization:

* **Step-1**
* **هضم**
* **Visrastyar**
* **Virastar**
* **Parsivar**: has pinglish converter
* **Persian pre-processing: preper**

ابزار برای Tokenization و Segmentation:

* **Step-1**
* **هضم**
* **Polyglot**
* **Tok-tok**
* **Segmental**
* **SeTPer:** regex based
* **Farsi-verb-tokenizer:** tokenize the Persian verbs
* **Parsivar**

برای تست Tokenizer word از جملات زیر استفاده شده است و خروجی هر ابزار مشخص شده است:

ابزار step-1 به ادرس:\*\*\* <http://step1.nlp.sbu.ac.ir/tokenizer/tokenizer.aspx>ابزار هضم: نسخه ۴.۰\*\*\*ابزار SeTPer به ادرس: http://stp.lingfil.uu.se/%7Emojgan/setper.html

|  |  |  |  |
| --- | --- | --- | --- |
| جمله | خروجی step1 | خروجی هضم | خروجی SeTPer |
| منبهدانشگاهرفتم!آیااورادیدم؟آیاخوردنیرااورد | منبهدانشگاهرفتم  !  آیا  او  را  دیدم  ؟  آیاخوردنیرااورد | ['منبهدانشگاهرفتم', '!', 'آیااورادیدم', '؟', 'آیاخوردنیرااورد'] | منبهدانشگاهرفتم  !  آیااورادیدم  ؟ آیاخوردنیرااورد |
| پدر بزرگ را می خواهیم به نا امید ترین شان معرفی کرده باشیم | پدربزرگ  رامی  خواهیم  به  ناامیدترین‌شان  معرفی  کرده باشیم | ['پدر', 'بزرگ', 'را', 'می', 'خواهیم\_به', 'نا', 'امید', 'ترین', 'شان', 'معرفی', 'کرده\_باشیم'] | پدر  بزرگ  را  می  خواهیم  به  نا  امید  ترین  شان  معرفی  کرده  باشیم |
| این داده را بایک جمله ی نهچندان سخت تست خواهیم کرد | این  داده  را  با  یک  جمله‌ی  نه‌چندان  سخت  تست  خواهیم کرد | ['این', 'داده', 'را', 'بایک', 'جمله', 'ی', 'نهچندان', 'سخت', 'تست', 'خواهیم\_کرد'] | این  داده  را  بایک  جمله  ی  نهچندان  سخت  تست  خواهیم  کرد |
| بعد يکروز سخت دريا نورد ي در لوله هاي فاضلاب | بعد  یکروز  سخت  دریانوردی  در  لوله‌های  فاضلاب | ['بعد', 'يکروز', 'سخت', 'دريا', 'نورد', 'ي', 'در', 'لوله', 'هاي', 'فاضلاب'] | مشابه هضم |
| تعدادروزها ی سال سیصدوشصتوپنج روز است و حدود ا ۲۴.۵ روز را یکچیزی حالا عدد ۵! خیلی خوب است. | تعداد  روزهای  سال  سیصدوشصتوپنج  روز  است  و  حدوداً  ۲۴  .  ۵  روز  را  یکچیزی  حالا  عدد  ۵  !  خیلی  خوب  است  . | ['تعدادروزها', 'ی', 'سال', 'سیصدوشصتوپنج', 'روز', 'است', 'و', 'حدود', 'ا', '۲۴', '.', '۵', 'روز', 'را', 'یکچیزی', 'حالا', 'عدد', '۵', '!', 'خیلی', 'خوب', 'است', '.'] | مشابه هضم |
| ۵\* ۲! + ۴ | ۵  \*  ۲  !  +  ۴ | ['۵\*۲', '!', '+۴'] | مشابه استپ‌وان |
| باران رحمت بی حسابش همه را رسیده و خوان نعمت بی دریغش همه جا کشیده. پرده ناموس بندگان به گناه فاحش ندرد و وظیفه روزی به خطای منکر نبرد. | باران  رحمت  بی  حسابش  همه  را  رسیده  و  خوان  نعمت  بی  دریغش  همه  جا  کشیده  .  پرده  ناموس  بندگان  به  گناه  فاحش  ندرد  و  وظیفه  روزی  به  خطای  منکر  نبرد  . | ['باران', 'رحمت', 'بی', 'حسابش', 'همه', 'را', 'رسیده', 'و', 'خوان', 'نعمت', 'بی', 'دریغش', 'همه', 'جا', 'کشیده', '.', 'پرده', 'ناموس', 'بندگان', 'به', 'گناه', 'فاحش', 'ندرد', 'و', 'وظیفه', 'روزی', 'به', 'خطای', 'منکر', 'نبرد', '.'] | مشابه هضم |

مشاهده میشود که ابزار هضم با استفاده از whitespace برای جداسازی کلمات استفاده میکند و در نتیجه با گذاشتن space اضافه براحتی دچار خطا میشود (و همچنین SeTPer) اما ابزار استپ‌وان اینگونه نیست. اما هر دو ابزار با بهم چسباندن کلمات دچار خطا میشوند.

Segmentation:

|  |  |  |
| --- | --- | --- |
| جمله | خروجی step1 | خروجی SeTPer |
| منبهدانشگاهرفتم!آیااورادیدم؟آیاخوردنیرااورد | منبهدانشگاهرفتم!  آیا او را دیدم؟  آیاخوردنیرااورد | منبهدانشگاهرفتم!آیااورادیدم؟آیاخوردنیرااورد |
| پدر بزرگ را می خواهیم به نا امید ترین شان معرفی کرده باشیم | پدربزرگ رامی خواهیم به ناامیدترین‌شان معرفی کرده باشیم | پدر بزرگ را می خواهیم به نا امید ترین شان معرفی کرده باشیم |
| این داده را بایک جمله ی نهچندان سخت تست خواهیم کرد | این داده را با یک جمله‌ی نه‌چندان سخت تست خواهیم کرد | این داده را بایک جمله ی نهچندان سخت تست خواهیم کرد |
| بعد يکروز سخت دريا نورد ي در لوله هاي فاضلاب | بعد یکروز سخت دریانوردی در لوله‌های فاضلاب | بعد يکروز سخت دريا نورد ي در لوله هاي فاضلاب |
| تعدادروزها ی سال سیصدوشصتوپنج روز است و حدود ا ۲۴.۵ روز را یکچیزی حالا عدد ۵! خیلی خوب است. | تعداد روزهای سال سیصدوشصتوپنج روز است و حدوداً ۲۴.۵ روز را یکچیزی حالا عدد ۵!  خیلی خوب است. | تعدادروزها ی سال سیصدوشصتوپنج روز است و حدود ا ۲۴.۵ روز را یکچیزی حالا عدد ۵!  خیلی خوب است. |
| ۵\* ۲! + ۴ | ۵\* ۲!  +۴ | ۵\* ۲!  + ۴ |
| باران رحمت بی حسابش همه را رسیده و خوان نعمت بی دریغش همه جا کشیده. پرده ناموس بندگان به گناه فاحش ندرد و وظیفه روزی به خطای منکر نبرد. | باران رحمت بی حسابش همه را رسیده و خوان نعمت بی دریغش همه جا کشیده.  پرده ناموس بندگان به گناه فاحش ندرد و وظیفه روزی به خطای منکر نبرد. | باران رحمت بی حسابش همه را رسیده و خوان نعمت بی دریغش همه جا کشیده.  پرده ناموس بندگان به گناه فاحش ندرد و وظیفه روزی به خطای منکر نبرد. |

برای ابزار استپ‌وان برای segmentation با استفاده از stopword جملات را از هم جدا میکند!

نرمالسازی در سطح حروف و فاصله: در ابزار هضم برای نرمالسازی space‌های اضافی را حذف میکند و بعضی از کلماتی که نیاز به نیم‌فاصله دارد را تصحیح میکند مثل ها و می و .... ظاهرا ابزار Parsivar بین نقاط نیز فاصله ایجاد میکند و همچنین نیم‌فاصله‌ها را نیز رعایت میکند.

|  |  |  |
| --- | --- | --- |
| جمله | هضم | Parsivar |
| به گزارش ایسنا سمینار شیمی آلی از امروز ۱۱ شهریور ۱۳۹۶ در دانشگاه علم و صنعت ایران آغاز به کار کرد. این سمینار تا ۱۳ شهریور ادامه می یابد | به گزارش ایسنا سمینار شیمی آلی از امروز ۱۱ شهریور ۱۳۹۶ در دانشگاه علم و صنعت ایران آغاز به کار کرد. این سمینار تا ۱۳ شهریور ادامه می‌یابد | به گزارش ایسنا سمینار شیمی آلی از امروز 11 شهریور 1396 در دانشگاه علم و صنعت ایران آغاز به کار کرد . این سمینار تا 13 شهریور ادامه می‌یابد |

جملات تست:

**تخت جمشید** ، مجموعه ای از کاخهای بسیار با شکوهی است که ساخت آنها در سال به 512 قبل از میلاد آغاز شد و اتمام آن 150 سال به طول انجامید. تختجمشید در محوطة وسیعی واقع شده که از یک طرف به کوه رحمت و از طرف دیگر به مرو دشت محدود است . این کاخهای عظیم سلطنتی در کنار شهر **پارسه** که یونانیان آن را **پرسپولیس** خوانده اند ساخته شده است.

تعداد محدودی از این جواهر ات در موزه ملی ایران نگهداری می شود. بزرگترین کاخ در مجموعه تخت جمشید کاخ مشهور به صد ستون است که احتمالا یکی از بزرگترین آثار معماری دوره هخامنشیان بوده و داریوش اول از آن به عنوان سالن بارعام خود استفاده می کرده است.

تایسون که اختر فیزیک ‌دان و مدیر آسمان ‌نمای هیدن در موزه‌ی تاریخ طبیعی آمریکا است، دوباره قرار است به نشنال جئوگرافیک باز گردد. طرفدارانژانر علمی‌ تخیلی کانال syfy را می‌شناسند. این کانال تا چند سال پیش با نام Scifi برنامه ‌هایش راپخش می‌کرد، اما بنا به گفته‌ی خودشان که تصمیم گرفته بودند مخاطب ان دیگر ی به جز علاقه‌ مندان علمی‌تخیلی راهم جذب کنند و برنامه‌های دیگری هم تولیدکنند. این برنامه اگرچه اولین تجربه‌ی برنامه شبانه‌ی علم‌ محور با چاشنی طنز و ادبیات عامه در تلویزیون به شمار می‌رود، اما می‌تواند آینده‌ی چنین برنامه‌هایی را مشخص کند و موفقیت یا عدم ‌موفقیت آن بیش از موفقیت یا شکست یک پروژه در بلندمدت بر حضور علم در نقش اصلی ومحوری دنیای رسانه‌های رسمی، تأثیرخواهدگذاشت.

اگر به دنیای علمی‌تخیلی و نقاطی که مرزهای آن با دنیای واقعی در هم ‌آمیخته می‌شود علاقه دارید، نام ایلان ماسک را در ذهن داشته باشید. کارآفرینی که اصالتاًاهل آفریقای جنوبی است و از او با نام مرد آهنی دنیای واقعی هم نام ‌برده می‌شود و در حال حاضر، مشغول ایفا ی نقش تونی استارک دنیای ما است.

اوهم‌ زمان ایده‌ی قطارهای فوق‌ سریع را مطرح کرده و اکنون به دنبال اینترنت رایگان جهانی است.

این سازمان فضایی تاکنون موفق به اثبات ۲۵۲۵ سیاره‌ی خارجی شده است که از این رقم، ده سیاره‌ی سنگی، هم‌ اندازه با زمین واحتمالا قابل سکونت هستند.

شاید برخی نگران باشند سیستم هوش مصنوعی گوگل  ستاره ‌شناسان راکنار بزند؛ اما به‌گفته‌ی ناسا جای نگرانی نیست.

چرا که ریاضی زبان فیزیک است، اما از آنجایی که پدیده‌های جهان ابتدا وجود داشته‌اند، مشاهده و درک شده‌اند و سپس برای آن‌ها مبانی ریاضیاتی تعریف شده است، با اندکی تلاش می‌توان مباحث مربوط به فیزیک کلاسیک را، بااجتناب از وارد شدن در مبانی ریاضی(که هدف تالیف این مجموعه است) توضیح داد. اما وارد شدن به دنیای فیزیک کوانتوم و پرهیز از مبانی ریاضی اجتناب ناپذیر است. چرا که دنیای فیزیک کوانتوم، به غایت انتزاعیو ذهنی است. همان‌طورکه شاید خودتان نیزبدانید، بحث‌ها و منازعات اصلی فیزیکدانان بزرگ نظیر؛ **فاینمن**، **اینشتین**، **بور** و ... همه و همه در مبانی کوانتوم بوده است و عامل تمامی این اتفاقات یک دلیل ساده است؛ فهم کامل فیزیک کوانتوم نا شدنی است!

در این قسمت از مجموعه، که به صورت اختصاصی به نا رسایی‌ های فیزیک کلاسیک و گام ‌های آغازین فیزیک کوانتوم می ‌پردازیم، استفاده از ریاضیات بیش از پیش حس می‌شد. چراکه درحالت کلی شما، تنها برای درک یک موضوع نیاز به دانستن صفرتاصد ریز موضوعات آن مطلب دارید، حال اگر بخواهید، نا رسایی موضوعی را پیدا کنید و برای حل آن راه کاری ارائه دهید، باید علاوه بر دانستن صفر تا صد ریز موضوعات بر تمامی مطالب مرتبط با آن موضوع نیز اشراف کامل داشته باشید، که مبانی ریاضی یکی ازاصلی ‌ترین ارکان آن است. با این وجود، تمامی تلاش بر این هدف گذاشته شده است، که مخاطبی که به علم علاقه‌مند است و در عین حال پایه‌ی ریاضیاتی قوی ندارد، بتواند مفاهیم و مباحث مطرح شده رادرک کند و از فهم آن‌ها لذت ببرد.

 فاجعه‌ی فرا بنفش یا فاجعه رایلی جینزاشاره به نتیجه‌ای دارد که ناشی از اصول فیزیک کلاسیک تقسیم مساوی انرژی و تابش نوسانگرهای بار دار برای توضیح تابش جسم سیاه در طول موج‌های کوتاه است.

فرمول پلانک هر چند موفقیت آمیز بود، اما نتیجه‌گیری ماهیت کوانتومی تابش از آن چندان الزامی نیست. بخش عمده‌ای در پذیرفتن آن از کار آلبرت اینشتین حاصل شد، که در سال ۱۹۰۵ با استفاده از مفهوم ماهیت کوانتومی نور، بعضی ازخاصیت‌های ویژه فلز ات را وقتی در معرض نور مرئی و فرا بنفش قرار می‌گیرند، توضیح داد.

کشف **اثر فوتو الکتریک** با کار **هاینریش هرتز** در سال ۱۸۸۷ آغاز شد.

 آزمایشی که سر راست ترین، مدرک ماهیت ذره‌ای تابش را در اختیار می‌گذارد، **اثر کامپتون** است. **آرتور هولی کامیون** کشف کرد، تابشی با یک طول موج معین (در ناحیه پرتو X ) که از یک ورقه‌ی فلزی می‌گذرد، به گونه‌ای پراکنده می‌شود که با نظريه‌ی تابش کلاسیک سازگار نیست. بنا به نظریه‌ی کلاسیک، ساز و کار این اثر عبارت است، از تابش مجدد نور توسط الکترون‌هایی که با تابش فرودی به نوسان وا داشته شده‌اند

**گریس هاپر (Grace Brewster Murray Hopper)**دانشمند آمریکایی علوم کامپیوتر بود که نام خودش را به‌عنوان یکی از اولین برنامه‌نویس‌های کامپیوتر مشهور **هاروارد مارک**جاودانه کرد. او یکی از پیشگامان برنامه‌نویسی مدرن بود و اولین کامپایلر برنامه‌نویسی را نیز توسعه داد. ایده‌ی زبان‌های برنامه‌نویسی مستقل از ماشین از ذهن این دانشمند زاده شد و بعد ها به خلق زبان برنامه ‌نویسی **کوبول** انجامید.

هاپر علاقه‌ی زیادی نیز به ارتش داشت و در دوران جنگ جهانی دوم برای ملحق شدن به ارتش آمریکا پیش‌قدم شد. البته او در ابتدا از پیوستن به ارتش منع شده اما در لیست نیروهای رزرو قرار گرفت. این بانوی دانشمند در توسعه‌ی کامپیوتر **UNIVAC I** نیز نقش داشته است. تلاش‌های هاپر برای توسعه‌ی کامپایلر در جهت این ایده بود که می‌توان به زبان انگلیسی با کامپیوترها صحبت کرد. دستاوردهای گریس هاپر و فعالیت او در ارتش باعث شده تا القاب متنوعی از **ملکه‌ی برنامه‌ نویسی** تا **گریس شگفت‌انگیز** به او داده شود. برخی از موشک‌ها و ابر کامپیوتر های آمریکایی نیز به نام این دانشمند ثبت شده‌اند. هاپر در طول مدت زندگی ۴۰ جایزه و درجه‌ی افتخاری متنوع دریافت کرد. حتی پس از مرگ هاپر نیز جایزه‌هایی به او اهدا شد که از این میان می‌توان **مدال آزادی** سال ۲۰۱۶ آمریکا را نام برد.

تعدادروزهایسالسیصدوشصتوپنجروزمیباشد. رومینادخترخیلیگلیاست. ایده‌یزبان‌هایبرنامه‌نویسیمستقلازماشینازذهنایندانشمندزادهشدوبعدهابهخلقزبانبرنامه‌نویسی**کوبول**انجامید. تابشفرودیبهنوسانواداشتهشده‌اند.کشفاثرفوتوالکتریکباکارهاینریشهرتزدرسال ۱۸۸۷ آغازشد.