بررسي سامانه جُهد

11 مرداد 1402

1. سند ارسالي مطالعه شد.
   1. معرفي سامانه‌هاي: جُهد، سحاب، قم‌نت و سنا.
2. دو سامانه جُهد و قم‌نت ارتباط مستقيمي با زيرساخت‌هاي هوش مصنوعي دارد.
3. قم‌نت ابزار نمايشي بسيار خوبي براي هستان‌ها و موجوديت‌هاي مفاهيم دارد.
4. اما؛
   1. مفاهيمي را در آن جستجو كردم.
   2. نتيجه قابل قبول نبود.
   3. بسيار دور از انتظار و آن‌چه توقّع مي‌رود.
5. سامانه جُهد را بررسي كردم و برخي مفاهيم علم فقه را جستجو كردم.
   1. اين سامانه منابع بسيار غني و قابل توجهي دارد:
      1. كتابخانه مدرسه فقاهت
      2. كتابخانه مدرسه اهل سنّت
      3. كتابخانه مدارك اسلامي
      4. تقريرات مدرسه فقاهت
      5. اَسناد نمايه شده از سحاب
6. ولي؛
   1. مطالبي كه انتظار مي‌رفت نمايش نداد.
   2. اطلاعاتي كه در كتابخانه مدرسه فقاهت هست.
   3. ولي در اختيار كاربر قرار نگرفت.
   4. پاسخ‌ها قابل قبول نبود.
7. پيشنهاد:
   1. امروزه كتابخانه‌هاي نرم‌افزاري بسياري توليد شده‌اند.
      1. هوش مصنوعي هم ابزارهاي فراواني دارد.
      2. بسياري از آن‌ها به صورت متن‌باز.
   2. اما؛
      1. آيا مي‌توانيم به همه اين ابزارها اطمينان كنيم؟
      2. آيا با مفاهيم زبان فارسي همان‌گونه خوب عمل خواهند كرد كه با مفاهيم ساير زبان‌ها؟
      3. آيا آن‌قدر كنترل روي اين ابزارها داريم كه بتوانيم شخصي‌سازي كنيم؟
   3. بالاخره؛
      1. ما قرار است سال‌هاي سال نرم‌افزارهاي هوش مصنوعي طراحي كنيم.
      2. مصرف‌كنندگان فراواني براي اين نرم‌افزارها در آينده يافت خواهد شد.
      3. اين موضوع به زودي و به اين سادگي دست از سر ما بر نمي‌دارد.
   4. ما منطق هوش مصنوعي را بلد هستيم.
   5. تمامي زيرساخت‌هاي نرم‌افزاري آن را هم داريم؛ زبان‌هاي برنامه‌نويسي مناسب.
   6. ما بايد؛
      1. خودمان از صفر ابزارهاي زيرساختي را طراحي كنيم.
      2. منطق هوش مصنوعي را براي زبان فارسي توسعه دهيم.
      3. منطق مزبور را تست كنيم و اصلاح و بهينه نماييم.
      4. بايد بتوانيم به تمام اين ابزارها مسلّط باشيم.
   7. ما نبايد؛
      1. از جستجوگرهايي مانند الاستيك سرچ استفاده كنيم.
      2. خود را محدود به ابزارهاي موجود كنيم.
      3. نبايد به بانك‌هاي مترادفات و مفاهيمي كه ديگران ساخته‌اند اكتفا نماييم.
   8. زيرا؛
      1. مي‌خواهيم بر بنيان‌هاي هوش مصنوعي سوار باشيم.
      2. و اين بنيان‌ها را اصلاح كنيم.
      3. مثل صنعت خودروسازي داخلي نشويم كه همه بنيان‌هايش خارجي‌ست!
8. به نظر مي‌رسد:
   1. شركت‌هاي توسعه‌يافته دنيا خودشان از پلت‌فرم‌هاي آماده استفاده نمي‌كنند.
   2. وقتي به كدهاي جيميل مثلاً نگاه كنيم، پلت‌فرم معروفي را در آن نمي‌بينيم.
   3. شركت‌هاي تأثيرگذار پلت‌فرم‌هاي خودشان را خودشان مي‌نويسند.
9. آن‌چه مي‌توانيم باشيم؛
   1. شايد شروع از صفر زمان بيشتري بگيرد.
   2. شايد مجبور شويم چرخ را دوباره اختراع كنيم.
   3. شايد خيلي از بانك‌هاي اطلاعاتي را بايستي خودمان پُر كنيم.
   4. خيلي از كتابخانه‌هاي نرم‌افزاري را بسازيم.
   5. اما؛
      1. رشد لگاريتمي خواهيم داشت؛ شتابي فراتر از سرعت معمول.
      2. در آينده «هر» كاري مي‌توانيم انجام دهيم.
      3. محدود به ابزارهاي ديگران نخواهيم بود.
      4. چشم به راه آپديت‌ها و نسخه‌هاي آپگريد نخواهيم بود.
      5. مجبور نخواهيم بود خواسته‌هاي خود را در حدّ ابزارهاي موجود فرو بكاهيم.
10. يك نقشه راه احتمالي؛
    1. متن‌ها را بر اساس واحد «مقاله» در نشريات و «فصل» در كتب تجزيه مي‌كنيم.
       1. همان كتابخانه‌هاي معرفي شده كه منابع‌شان در اختيار است.
    2. تمامي حروف را حذف مي‌كنيم.
       1. حروف اضافه، ربط، نشانه و هر چه غير از اسم و فعل است.
       2. افعال اِسنادي هم حذف مي‌شود؛ است، بود، گشت، شد.
    3. افعال را تبديل به مصدر مي‌كنيم؛ رفتيم تبديل مي‌شود به رفتن.
    4. يك بانك اطلاعات خواهيم داشت از اسم‌ها و مصدرها.
    5. تمام اسم‌ها و مصدرهاي موجود در يك واحد اطلاعاتي با هم مرتبط مي‌شوند.
       1. اگر يك واحد اطلاعاتي هزار كلمه داشته باشد، هر كلمه با 999 كلمه مرتبط مي‌شود.
       2. هر ربط يك ارزش و رتبه دارد.
       3. رتبه و ارزش براي كلمات در يك مقاله عدد «يك» خواهد بود.
    6. اين ربط به واحد اطلاعات «مقاله» يا «فصل» متصل خواهد ماند.
    7. اكنون واحد اطلاعاتي را پاراگراف يا بند در نظر مي‌گيريم و تمام واحدهاي قبلي را تجزيه مي‌كنيم.
    8. ارتباط اسم‌ها و مصدرهاي واحدهاي جديد يك رتبه جديد مي‌گيرند.
       1. رتبه «يك» به رتبه فعلي آن‌ها اضافه مي‌شود.
       2. اين واژه‌ها و مفاهيم رتبه «دو» در ارتباط خواهند داشت.
       3. اين رتبه‌ها به بند مذكور متصل خواهد ماند.
    9. اكنون هر بند را بر حسب جمله تجزيه مي‌كنيم، بر اساس وجود نقطه در انتهاي هر جمله.
       1. در تحليل جمله مي‌شود از «معناشناسي هوشمند» بهره برد.
       2. ولي در فاز اول نيازي اصلاً به اين تحليل‌هاي معنايي نداريم.
    10. دوباره ارتباط‌ها را بررسي كرده و رتبه‌هاي آن‌ها را به «سه» ارتقا مي‌دهيم.
    11. تجزيه بعدي بر اساس ملاحظه مجاورت واژگان خواهد بود.
    12. هر واژه با واژه بعدي خود مرتبط محسوب مي‌شود؛ اهل كتاب: اهل + كتاب.
    13. ارتباط ميان واژه‌هاي متوالي يك ارتقا رتبه خواهد داشت.
    14. حالا؛
        1. يك بانك اطلاعات از مفاهيم داريم؛ اسم و مصدر.
        2. يك بانك اطلاعات ديگر از روابط ميان اين مفاهيم.
        3. هر رابطه يك رتبه عددي دارد.
        4. اگر كاربر عبارت «اهل كتاب» را جستجو كند:
           1. رابطه اهل و كتاب را با هم ملاحظه مي‌كنيم.
           2. بالاترين رتبه را در كدام بند از كدام مقاله دارد؟
           3. مفاهيمي را كه بالاترين رتبه ارتباطي با اهل و بالاترين رتبه ارتباطي با كتاب دارند فهرست مي‌كنيم.
           4. آن‌ها در كدام بند از كدام مقاله بيشتري رتبه ارتباطي را دارند؟
           5. ما بهترين معرفي و توضيح درباره «اهل كتاب» را پيدا كرده‌ايم.
        5. اين ساده‌ترين ابزار جستجوي هوشمند در كتابخانه‌هاست كه در دسترس است.
    15. آن را با بانك‌هايي از مترادفات، متضادها و ساير روابط اسامي در زبان فارسي تقويت مي‌كنيم.
        1. اهل كتاب = يهودي = نصراني = مجوس = …
    16. در گام‌هاي بعد به سمت توليد ابزارهايي مي‌رويم كه پاراگراف‌هاي به دست آمده را تحليل كند.
    17. آن‌ها را خلاصه و چكيده كند.
    18. مفهوم اصلي را بيرون بكشد.
    19. در قالب جملات فارسي عاميانه و مخاطب‌محور ارائه كند.
11. به اين ترتيب ما مي‌توانيم جستجوگر هوشمند گفتگومحور خودمان را داشته باشيم.
12. اين يك پيشنهاد اوليه و خام است و نياز به گفتگوها و بررسي‌ها و نقد و نظرهاي بيشتر دارد.
    1. ولي قابل حصول است.
    2. مالِ خودمان است.
    3. مي‌توانيم در آينده با دست باز آن را ارتقا دهيم.

سيدمهدي موسوي موشَّح - قم المقدسة