



دانشکده مهندسی و علوم کامپیوتر

## حل جدول کلمات متقاطع با کمک تکنیک های هوش مصنوعی

پروژه درس هوش مصنوعی و سیستم های خبره

استاد: دکتر شمس فرد

پاییز ۱۴۰۱

## مقدمه

در این پروژه قصد داریم یک نسخه ی ساده سازی شده از جدول کلمات متقاطع را به کمک تکنیک های هوش مصنوعی حل کنیم.

جدول کلمات متقاطع را احتمالاً در مجلات و روزنامه های بسیار دیده اید. در این جداول تعدادی شرح (یا سوال) داریم و برای هر شرح یک جواب وجود دارد که یک کلمه ی خاص است. ممکن است یک حرف از یک کلمه در یک ستون، یک حرف از یک کلمه در یک سطر نیز باشد و شما باید جدول را طوری بچینید که این موارد همخوانی داشته باشند. در شکل زیر یک نمونه از این جدول را می بینید:

۹	۸	۷	۶	۵	۴	۳	۲	۱	
				■					۱
				■					۲
						■			۳
			■						۴
■	■						■	■	۵
					■				۶
		■							۷
				■					۸
				■					۹

یک نمونه شرح برای این جدول به اینصورت است:

1- افقی: جنس و متاع - جمع ادیب

1- عمودی: تفحص و جستجو - گوشت بریان روی آتش

باید کلمه ی مناسب را برای این دو شرح انتخاب کنیم بطوری که حرف اول ردیف 1

افقی (سبز) با حرف اول ستون 1 عمودی (قرمز) یکی شود، و به همین ترتیب بقیه

جدول را تکمیل می کنیم

شرح های داده شده می توانند دو نوع باشند:

1) یک کلمه ای که برای این حالت باید مترادفی مناسب با کلمه ی داده شده پیدا کنیم، مثلاً «دلداری» یا «بشارت دادن» یا حتی

«جنس و متاع» هم جزو شرح های یک کلمه ای است.

2) چند کلمه ای، مانند «از اقوام ایرانی» یا «گوشت بریان روی آتش» که برای حل آنها نمی توان مترادف پیدا کرد و باید از روش دیگری استفاده کرد.

## شرح پروژه

در این پروژه انتظار می رود به کمک هوش مصنوعی بتوانیم جدول کلمات متقاطع را حل کنیم، اما جدول کلمات خواسته شده در این پروژه یک نسخه ی ساده سازی شده از مسئله ی اصلی است و تفاوت هایی با نسخه ی اصلی دارد که در ادامه بیان می شوند.

- شرح های **یک کلمه ای** مشابه جدول اصلی هستند و برای این قسمت شما باید ابتدا برای هر شرح فهرستی از مترادفات (یا متضادها) آنرا استخراج کنید، برای استخراج این مترادفات استفاده از هر روشی مجاز است و انتخاب و پیاده سازی بر عهده ی خود شماست.

• شرح های چند کلمه ای با نسخه ی اصلی مسئله به منظور ساده سازی تفاوت هایی دارند. در این قسمت سوالات از دسته های محدودی پرسیده می شود مثلاً اگر فرض کنیم فقط دو دسته ی «خودرو» و «غذا» را داشته باشیم، تعدادی از شرح های محتمل می توانند این ها باشند: «غذایی تهیه شده از گوشت چرخ شده»، «خودرویی رایج در ایران»، «غذایی مخصوص جنوب ایران» و ...

برای این قسمت شما نیاز است که ابتدا برای هر دسته ی محتمل (که تعداد محدودی خواهد بود)، تمام جواب های محتمل را از پیش جمع آوری کنید و در جایی نگه داری کنید. سپس در حین حل جدول برای شرح های چند کلمه ای باید بتوانید دسته ی هر کدام را تشخیص دهید، سپس با توجه به شرایط جدول و تعداد کلمات مورد نیاز با کمک روش های هوش مصنوعی کلمه ی مناسب را در جای خود بنشانید.

**توجه:** منظور از تمام جواب های محتمل جوابهاییست که ممکن است در جدول استفاده شده باشند، مثلاً یک حیوان بسیار نادر که کمتر کسی اسم آنرا شنیده قطعاً جواب نخواهد بود. البته واضح است که نادر بودن یا نبودن یک جواب نسبی است و بدیهی است که هر چه جواب های محتمل بیشتری پوشش داده شوند، عملکرد سیستم شما بهتر خواهد بود.

دسته های مورد نظر در ادامه گفته خواهند شد.

## بازنمایی مسئله

ورودی یک جدول  $n*m$  است که  $n$  تعداد سطرها و  $m$  تعداد ستونهاست. این جدول بصورت یک دنباله  $n*m$  از 0 و 1 نمایش داده می شود. صفرها خانه های خالی و 1 ها خانه های سیاه هستند. مثلاً جدول زیر به شکل دنباله داده شده نمایش داده می شود (جدول فارسی است و راست به چپ خوانده می شود):

	01000
	00100
	00001
000010010001000	

علاوه بر شکل جدول در ورودی تعدادی شرح نیز داده می شود که به ترتیب شرح کلماتی است که باید در جدول قرار گیرند. بین شرح هر کلمه با بعدی در همان سطر یا ستون یک # گذاشته می شود و بین شرح کلمات هر سطر (یا ستون) با بعدی یک @ می آید. اول و آخر کل شرح ها هم با & مشخص می شود و فرض بر این است که سه علامت @ و # و & در متن شرح ها دیده نخواهند شد.

شرح ها به ترتیب افقی و سپس عمودی نوشته می شوند مثلاً بازنمایی شرح زیر برای جدول فوق به شکل داده شده خواهد بود:

افقی:

- 1- سلطان جنگل
- 2- ضمیر اول شخص مفرد- بله به انگلیسی
- 3- نام پیامبری که برادران به چاه افکندند

عمودی:

- 1- دوم شخص جمع بدون انتها!

2- "یعنی" در نوشتار محاوره ای!

3- -

4- نام سوره ای از قران مجید

5- دریغ و افسوس

نمایش شرح در ورودی:

&سلطان جنگل @ ضمیر اول شخص مفرد # بله به انگلیسی @ نام پیامبری که برادران به چاه افکندند @ دوم شخص جمع بدون انتها! @ یعنی در نوشتار محاوره ای! @-@ نام سوره ای از قران مجید @ دریغ و افسوس &

خروجی دنباله ای با سایز  $n*m$  از حروف است که بجای خانه های سیاه @ گذاشته می شود مثلا حل جدول بالا به این شکل نمایش داده میشود:

&شیر#@من#@پیس#@#یوسف@شم#@پنی@-#@#پیس@اسف&

(البته دنباله در اینجا راست به چپ نوشته شده و در واقعیت ش در خانه اول دنباله قرار دارد)

جدول حل شده با توجه به دنباله ی خروجی:

ش	ی	ر		ا
م	ن		ی	س
	ی	و	س	ف

پس از بدست آوردن دنباله ی خروجی باید جدول را از روی آن بازسازی کنید و بصورت گرافیکی نمایش دهید.

## دسته بندی

برای آموزش دسته بند نیاز است داده ی آموزشی دارای برچسب داشته باشیم. به منظور افزایش اندازه ی داده، هر کدام از دانشجویان باید برای هرکدام از دسته ها تعدادی شرح چند جمله ای طراحی کند و کتگوری آنرا مشخص کند و آنها را در یک مخزن عمومی برای استفاده سایر دانشجویان قرار دهد. انجام این قسمت بخشی از نمره ی پروژه خواهد بود. جزییئات لازم برای این بخش در آینده در اختیارتان قرار خواهد گرفت.

## فازبندی

با توجه به اینکه پروژه از چندین مازول کوچکتر تشکیل شده، برای سهولت و یکپارچگی روند انجام پروژه توصیه میشود از فازبندی زیر استفاده کنید:

**فاز اول)** دریافت ورودی و نمایش خروجی در قالب جدول، در این مرحله با فرض داشتن جواب ها باید جدول خروجی را باز تولید کنید.

**فاز دوم)** در این مرحله با فرض داشتن کلمات محتمل برای هر شرح، باید راه حل های محتمل را به کمک الگوریتم های مناسب بیابید.

**فاز سوم)** در این مرحله کار اصلی حل جدول یعنی یافتن پاسخ ها انجام میشود، برای شرح های تک کلمه ای مترادف (یا متضاد) و برای شرح های چند کلمه ای باید دسته ی جواب را تشخیص دهید.

**فاز نهایی)** ترکیب تمام فاز های پیشین، در **فاز سوم** شما توانسته اید برای شرح های یک کلمه ای لیستی از مترادفات محتمل و برای شرح های چند کلمه ای لیستی از جواب های محتمل با توجه به دسته ی تشخیص داده شده استخراج کنید. حال شما باید به کمک الگوریتمی که در **فاز دوم** پیاده سازی کرده اید، جواب های جدول را از بین کلمات لیست های محتمل بدست بیاوردی و در نهایت به کمک آنچه در **فاز اول** انجام دادید نمایش دهید.

### تولید داده های مورد نیاز)

- تعدادی جدول به همراه شرح و حل آنها در اختیارتان قرار خواهد گرفت و می توانید از آنها برای پیاده سازی و تست استفاده کنید.
- دسته ی دوم داده های مورد نیاز جفت های (شرح و دسته) است که به طور خاص برای آموزش دسته بند مورد استفاده قرار میگیرند، برای این قسمت بازه زمانی مشخصی تعیین میشود و همه ی دانشجویان بصورت موازی با انجام پروژه خود باید این داده ها را تهیه کنند و در محل مشخص شده برای استفاده ی بقیه دوستان خود قرار دهند. داده های تولید شده در این بخش تشکیل دهنده ی بخشی از نمره ی شما هستند و همچنین برای فاز های سوم و چهارم مورد نیاز هستند بنابراین مهلت تعیین شده برای آن قابل تمدید نخواهد بود.

### دسته ها

1. حیوانات
2. فصول و ماه های سال (میلادی و شمسی) (مثلاً: دسامبر، بهار، مهر و ...)
3. اشیا
4. غذاها (و نوشیدنی)
5. کشورها و شهرهای جهان
6. مشاغل (مثلاً: نجار، قصاب و ...)
7. برندها و مدل های خودرو (مثلاً: مزدا، پراید، تویوتا و ..)

8. میوه ها
9. سلسله ها و پادشاهان حکومت های ایران (مثلاً: ساسانیان، قاجار و .. یا خشایارشا و ...)
10. اصطلاحات مربوط به فیلم (مثلاً ژانر های فیلم ها، فیلم ها معروف و ..)
11. عناصر جدول تناوبی
12. رنگ ها
13. ضمائر فارسی
14. سوره های قرآن
15. اسامی پیامبران
16. اسامی دانشمندان معروف (مثلاً: کوری، اینشتین، رادرفورد و ...)
17. صفت های فارسی (مثلاً: جاهل، مغرور و ...)
18. ورزشکاران جهان (مثلاً: مسی، کوبی و ...)
19. تیم های فوتبال جهان
20. مواد غذایی
21. شاعران و نویسندگان (مثلاً: هوگو، تولستوی، سعدی و ...)
22. اسامی فارسی
23. انواع گل ها و گیاهان
24. مکان های طبیعی و توریستی جهان (مثلاً دماوند، بیستون، ...)
25. اعضای بدن
26. اصطلاحات پزشکی (مثلاً اسم بیماری ها و ...)
27. آثار ادبی و هنری (مثلاً: بوستان، مونا لیزا، ...)
28. زبان ها
29. کلمات ضرب المثل ها (مثلاً در ضرب المثل «آسیاب به نوبت» کلمات «آسیاب» و «نوبت»)
30. سایر

- برای درک بهتر شکل و فرمت کلماتی که باید در لیست ها قرار دهید توصیه می شود و نمونه ها و مثال هایی از جواب های جدول توجه کنید.
- اگر دسته بند شما یک شرح را متعلق به دسته ی «سایر» تشخیص دهد آن خانه خالی می ماند، برای حل آنها می توانید از هر روش دلخواهی استفاده کنید (فقط به شرطی که توسط دسته بند «سایر» تشخیص داده شده باشد)، یک روش پیشنهادی می تواند تهیه لیستی از جواب های رایج جداول یا جستجو در سایت هایی است که سوالات به جواب های آماده جدول کلمات را دارند (مثلاً سایت آبادیس). در بسیاری موارد هم این خانه ها از طریق حل بقیه ی جدول بصورت خودکار پر خواهند شد.
- در برخی موارد لیست های تهیه شده اشتراکاتی هم دارند، مثلاً گل «لاله» و اسم «لاله»