



درباره

ویژگی ها

كاربردها

مشخصات فني

نمونه مفاهيم

مشتریان ما

ارتباط با ما

# َٰٰ≡َ] فهرست مطالب

### [i] درباره الفبا

#### به نام خداوند جان و خرد

نویسهخوان ابزاری است که نوشتهها را از دل تصاویر متنی بیرون میکشد و کاغذ را وارد دنیای رایانه میکند. در دهه اخیر، چند ابزار نویسهخوان فارسی آماده شدهاند؛ اما کیفیت آنها نیازهای مخاطب را به خوبی برطرف نکرده است. حالا گروه روشن با استفاده از فناوری روز دنیا، ابزار «الفبا» را آماده کرده است تا اسناد را سریع و دقیق تحلیل کند.

گروه **روشن** از سال ۱۳۹۰ با همت جمعی از دانشجویان دانشگاه علم و صنعت ایران شروع به کار نموده است. این گروه هدف خود را تلاش در راستای تحقق تمدن نوین اسلامی میداند و در قالب شرکت دانشبنیان راهکار پردازش ژرف برای پاسخ به نیازهای کشور، محصولات نرمافزاری ارائه میکند.

الفبا، به دنبال ارائه یک نویسهخوان فارسی کامل و جامع است.

## َ `√ ً وَيژگَی های الفبا

مختلف

آن استفاده شده باشد.



خوش می روی بر نام ما

ای درشکسته جام ما

در صفحات متن معمولا عبارات بر اساس اهمیتشان با

قلم و اندازه متفاوت نوشته میشوند. الفبا صفحه را

میخواند درحالی که چند قلم با اندازههای گوناگون در



### خواندن نویسههای فارسی و انگلیسی

Laws that have

endured for centuries

لابهلای متون فارسی، عبارات انگلیسی زیاد دیده میشوند. الفبا هر دو زبان را در کنار هم میخواند و نویسهها را با توجه به متفاوت بودن جهت نوشتار آنطور میچیند که متن به صورت یکپارچه خوانده شود. همچنین خواندن متون به زبانهای عربی و دری نیز برای الفبا میسر است.

# خواندن متن با قلمها و اندازههای

یا برجا مانده است

### تحليل ساختار سند

کلمههای فارسی در کنار هم میآیند و خطوط را شکل میدهند. خطهای پیاپی هم، پاراگراف را میسازند. این مفاهیم که در نوشتهها واضحاند، در تصویر سند از دست میروند و تبدیل به فاصلهها و طولها میشوند. الفبا همین فاصلهها را تحلیل میکند و ساختار سند، از جمله خطوط و پاراگرافها را پیدا میکند. همچنین محل عکسهای سند در خروجی حاصل از تحلیل آن مشخص مىشود.

# َ ویژگی های الفبا (ادامه)

قُولُوا لِلنَّاسِ حُسْناً

← قولوا للناس حسنا

شوریدهای ست که دیوانگان به آفتابش سیردهاند که دیوانگان به آفتابش سیردهاند

بنی آدم اعضای یک پیکرند

بنی آدم اعضای یک پیکرند

#### خواندن نویسهها در متون با اعراب

آیات یا احادیث در میان نوشته، با اعراب ظاهر میشوند و حرکتگذاری برای عباراتِ فارسی هم چندان عجیب نیست. الفبا این نوشتهها را میخواند و حرکتها را حذف میکند.

### اصلاح چرخش و نویز تصویر سند

گاهی تصویر اسکنشده متن، با نویز همراه است یا ممکن است کاغذ آن کمی چرخیده باشد. این اشکالات پیش از تحلیل متن سند، برطرف میشوند.

### قابلیت آموزش برای متون دستنویس

ابزار نویسهخوان روی قلمهای تایپی آموزش دیده است اما آنها را به خاطر نسپرده. یعنی حتی میتواند قلمهای جدید را هم بخواند؛ اگرچه در این کار اشتباه هم میکند. با این توضیح، همین ابزار برای خواندن متون دستنویس هم قابل آموزش است و نسخه کنونی آن، بعضی اعداد دستنویس را درست میخواند.





#### حفظ ساختار دادههای درون جدول

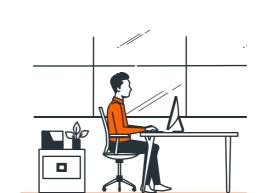
سندهایی که شکلی از پژوهش یا تحلیل را ارائه میکنند، پر هستند از جدول و میتوان گفت که اعداد یا نوشتههای درون این جدولها اهمیت مضاعفی دارند. الفبا با تشخیص ساختار جدول و مشخص کردن خانههای آن، محتوی هر خانه را با همان ساختاری که در جدول قرار گرفته، باز میگرداند.

### مشخص کردن خطاهای احتمالی

در سامانههای مبتنی بر «هوش ماشینی» رسیدن به محصول بدون خطا، تقریبا غیر ممکن است. الفبا هم گاهی نوشتهها را اشتباه میخواند. اما معمولا از اینکه ممکن است اشتباه خوانده باشد، آگاه است و با مشخص کردن میزان اطمینان از خواندن هر بخش، به استفاده کننده میگوید که در مورد آن بخش از نوشته «شک» دارد.

نویسهخوان الفبا، در جاهای مختلفی همچون کسبوکارها، سازمانها و کتابخانهها مفید و راهگشاست.

# [⊚] كاربردهاي الفبا



#### در کسبوکارها

نرمافزارهای زیادی با کتاب، مقاله، روزنامه، مجله، فاکتور، صورتحساب، فکس و به طور خلاصه کاغذ کار میکنند که محتوی آن به دلیل PDF سروکار دارند. بعضی از این ابزارهای خدمترسان هم با فایل اشکالات قدیمی نرمافزارها، در دسترس نیست. این اسناد اگرچه به ظاهر حروفچینی شدهاند، یک الفبا برای خوانده شدن نیاز دارند.



انبوه پروندهها و نامهها در اتاقهای بایگانی خاک میخورند و منتظر نویسهخوان فارسی هستند. عنوان «دولت الکترونیک» که رسانهها زیاد گفتهاند و زیاد شنیدهایم، چنین پیشنیازی دارد و قبل از وقوع آن بیگمان باید سالها پرونده غیرالکترونیک، نویسهخوانی شوند.



#### در کتابخانه

کتابخانههای پیشگام، در سالهای اخیر به سراغ اسکن کردن صفحات کتابها رفتهاند. این حجم عظیم از صفحات اسکنشده که تا کنون تنها جنبه نگهداری کتابها در قالب الکترونیکی را داشتهاند، با استفاده از ابزار نویسهخوان میتوانند قابل جستجو و بازیابی شوند.



نویسهخوان الفبا از نظر مشخصات فنی کامل است

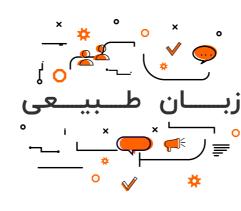
## رُّهِ ً مشخصات فني الفبا





#### فناوری یادگیری عمیق

زیربنای الفبا استفاده از فناوری یادگیری عمیق برای خواندن متن است. این فناوری در سالهای اخیر، دنیای هوش مصنوعی را متحول کرده و هماکنون بسیاری از غولهای عرصه فناوری اطلاعات، سرمایهگذاریهای عظیمی در این حوزه انجام دادهاند. استفاده از یادگیری عمیق است که باعث میشود این نویسهخوان بتواند قلمهای جدید را هم بخواند.



### پشتیبانی از پردازنده گرافیکی

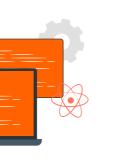
فناوری یادگیری عمیق، حجم پردازش قابلتوجهی نیاز دارد و اصلا همین بالا رفتن توان محاسباتی بوده که تحقیقات دراین حوزه را ممکن کرده است. پردازنده گرافیکی که برای موازیسازی محاسبات طراحی شده است، به الفبا امکان میدهد تا اسناد را چند برابر سریعتر از حالتی که تنها به پردازنده معمولی دسترسی دارد، تحلیل کند. این پردازندهها (کارتهای گرافیک) هماکنون نسبت به توان پردازشی بسیار بالایی که فراهم میکنند، ارزان قیمت و سهلالوصول هستند.

### پردازش زبان طبیعی

الفبا جملههای فارسی را میشناسد. یعنی گاهی وقتها که میان دو شکل مختلف یک عبارت تردید دارد، هر دو را در میان دیگر کلمات خط نوشته قرار میدهد، تا ببیند کدامیک جمله را بهتر کامل میکند. ابزاری که این کار را انجام میدهد «مدل زبانی» نام دارد و از یادگیری عمیق برای آموختن زبان استفاده میکند.

### هُ مشخصات فني الفبا (ادامه) [ه







در مواجهه با انبوه اسناد، بهرهبردن از پردازش توزیعشده به نویسهخوان این امکان را میدهد که از هر تعداد «ماشین» و «کارت گرافیک» فراهمشده، استفاده کند و با مدیریت مناسب منابع، به نیازمندیهای «بزرگ مقیاس» پاسخ دهد.

#### واسط برنامهنویس کامل

سامانه نویسهخوان برای استفاده دیگر ابزارها آماده است فراهم JSON یعنی واسط مناسبی در قالب تقاضاهای میکند و میتواند به آسانی با سامانههای دیگر هماهنگ شود.

#### دقیق و سریع

کشف یک ابزار یادگیرنده است و دقت آن وابسته است به دادههایی که روی آنها آموزش داده شده. به این ترتیب، با افزایش دادههای درست، به آسانی دقت تشخیص آن بالا میرود. کشف به صورت مداوم دقت را اندازه میگیرد و به کاربر نمایش میدهد تا جمعآوری دادههای مهم تسریع شود. از نظر عملکرد هم، کشف میتواند در هر ثانیه ۱۰ هزار واژه را با سخت افزار مناسب (کارت گرافیک) تحلیل کند و این عددها در تعداد ماشینهای در اختیار سامانه، ضرب میشوند.

#### متن خروجي الفبا

### مرد صحرا

اه ای صبور ساده دل ای مرد صرا ای درنگاهت آفتاب و صبح و دریا در چشمهایت می توان خورشید را دید ای وارث شبهای توفان خیز یلدا بامن بگو این فصل، فصل چندمین است گزباغ وبارآن دوری ای مرد تماشا مرغ گلویت پرنمی گیرد به باغی چشمت نمی بیند بجز بیحاصلی ها ایکاش میشد بارها بوسید و بویید آن دستهای تشنهی زخم آشنا را

### [�] نمونه مفاهيم الفبا

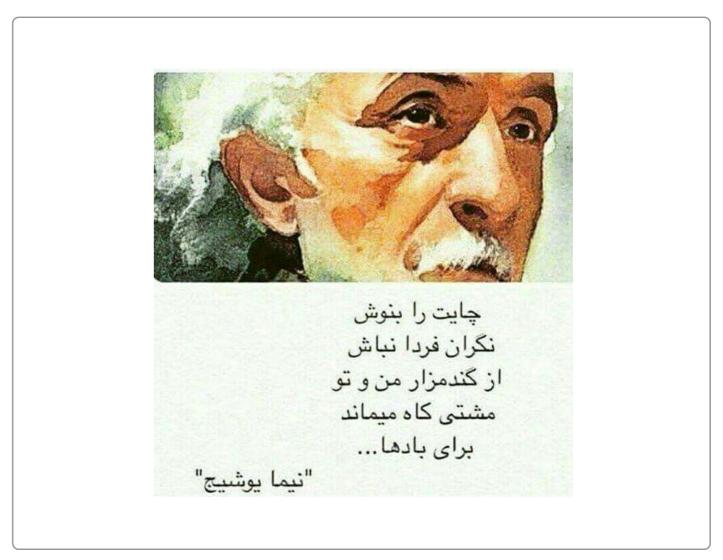


# مرد صحرا

آه ای صبور ساده دل ای مرد صحــرا ای درنگساهت آفتاب و صبسح و دریسا در چشمهایت میتوان خورشید را دیـد ای وارث شبسهای تسوفان خیسز یسلدا بامن بگو این فصل، فصلچندمیناست کـزباغ وبــاران دوری ای مــرد تــماشا مرغ گلویت پسرنمی کسیرد بسه بساغی چشمت نمیبیند بجز بیحاصلیها ایکاش می شد بارها بسوسید و بسویید آن دستهسای تسسنهی زخم آشنسا را

تصوير ورودي

# [�] نمونه مفاهيم الفبا



تصوير ورودي

#### متن خروجي الفبا



#### متن خروجي الفبا



# [ط] نمونه مفاهيم الفبا



		ازنامه	تر		
1490	ماه	اسفند	4+	تاريخ	,3

1895/17/49	1890/17/80	داراييها
ميليون ريال	میلیون ریال	
117:140 A7:747 A7:040:1 P4:444 A3:434	74.+70 94.14 1.490.41A 74.515 733.470	دار اییهای جاری: موجودی نقد سرمایه گذاریهای کوتاه مدت دریافتنی های تجاری و غیر تجاری مبلغ قابل بازیافت پیمانها موجودی مواد و کالا پیش پرداختها
732,577.7	T.+78.778	مع داراییهای جاری

تصوير ورودي

#### متن خروجي الفبا

#### Chapter 6

#### FINAL REVIEW

First, however, let us summarize these themes and variations You have earned the basic elements of checkmate, the back-rank mates and defenses, dis placing themes, and an assortment of sacrificial com binations The following diagrams illustrate these concepts in their basic form

TURN THE PAGE AND CONTINUE WITH FRAME 240

### [�] نمونه مفاهيم الفبا



#### Chapter 6

#### FINAL REVIEW

First, however, let us summarize these themes and variations. You have learned the basic elements of checkmate, the back-rank mates and defenses, displacing themes, and an assortment of sacrificial combinations. The following diagrams illustrate these concepts in their basic form.

TURN THE PAGE AND CONTINUE WITH FRAME 240

تصوير ورودي

















CIVILICA







(ل) جمهوری اسلامی ایران وزارت کثور



# [۵] مشتریان ما

[ٰٰٰٰٰٰٰٰٰ الرتباط با ما

شرکت دانشبنیان راهکار پردازش ژرف

alefba@roshan-ai.ir

4°44°016 - 14°

تهران، نارمک خیابان فرجام، خیابان حسینعلی، پلاک ۹، واحد ۲۴