

بِسْمِ اللَّهِ الرَّحْمَنِ الرَّحِيمِ



درآمدی بر زبان‌شناسی رایانشی در زبان عربی

دکتر منصور بن محمد الغامدی؛ دکتر عبدالعزیز بن عبدالله المهبوبی
احمد رومی محمد عبدالرحمن؛ دکتر اشراق علی احمد الرفاعی
دکتر صلاح راشد الناجم؛ دکتر ولید بن عبدالله الصانع

ترجمه:

دکتر سکینه محمدی
دکتر امیر مقدم متقی
دانشیار دانشگاه فردوسی مشهد
دکتر محسن کاهانی
استاد دانشگاه فردوسی مشهد
دکتر مهسا جدیدالاسلامی

عنوان قراردادی:
عنوان و نام پدیدآور:

مدخل الى اللسانيات الحاسوبية. فارسي
درآمدی بر زبان‌شناسی رایانشی در زبان عربی / منصور بن محمد الغامدی... [و دیگران]؛ ترجمه
سکینه محمدی... [و دیگران]؛ ویراستار علمی عبدالاحد غیبی؛ ویراستار ادبی هانیه اسدیپور
فعال مشهد.

مشخصات نشر:
مشخصات ظاهری:
فروست:

مشهد: دانشگاه فردوسی مشهد، انتشارات، ۱۴۰۲.
۱۶۴ ص.
انتشارات دانشگاه فردوسی مشهد؛ ۸۹۰.

شابک:
یادداشت:

ISBN: 978-964-386-584-9

منصور بن محمد الغامدی، عبدالعزیز بن عبدالله المهبوبی، احمد رومی محمد عبدالرحمن، اشراق
علی احمد الرفاعی، صلاح راشد الناجم، ولید بن عبدالله الصانع.
ترجمه سکینه محمدی، امیر مقدم متقی، محسن کاهانی، مهسا جدیدالاسلامی.
کتابنامه. نمایه.

یادداشت:
یادداشت:
موضوع:

زبان‌شناسی کامپیوتری -- کشورهای عربی
زبان عربی -- داده‌پردازی

الغامدی، منصور بن محمد، ۱۹۵۶-م.
محمدی، سکینه، ۱۳۵۸- مترجم
غیبی، عبدالاحد، ۱۳۵۸-، ویراستار
دانشگاه فردوسی مشهد، انتشارات.
P۹۸

شناسه افزوده:
شناسه افزوده:
شناسه افزوده:
شناسه افزوده:
رده‌بندی کنگره:

رده‌بندی دیویی:
شماره کتابشناسی ملی:

۴۱۰/۲۸۵
۹۳۴۶۶۳۲

درآمدی بر زبان‌شناسی رایانشی در زبان عربی

پدیدآورندگان: دکتر منصور بن محمد الغامدی؛ دکتر عبدالعزیز بن عبدالله المهبوبی
احمد رومی محمد عبدالرحمن؛ دکتر اشراق علی احمد الرفاعی
دکتر صلاح راشد الناجم؛ دکتر ولید بن عبدالله الصانع
ترجمه: دکتر سکینه محمدی؛ دکتر امیر مقدم متقی
دکتر محسن کاهانی؛ دکتر مهسا جدیدالاسلامی

ویراستار علمی: دکتر عبدالاحد غیبی

ویراستار ادبی: هانیه اسدیپور فعال مشهد

مشخصات: وزیری، ۱۵۰ نسخه، چاپ اول، پاییز ۱۴۰۲

چاپ و صحافی: چاپخانه دقت

بها: ۱/۲۵۰/۰۰۰ ریال

حق چاپ برای انتشارات دانشگاه فردوسی مشهد محفوظ است.



انتشارات
۸۹۰

مراکز پخش:

فروشگاه و نمایشگاه کتاب پردیس: مشهد، میدان آزادی، دانشگاه فردوسی مشهد، جنب سلف یاس

تلفن: ۳۸۸۰۲۶۶۶ - ۳۸۸۳۳۷۲۷ (۰۵۱)

تهران، میدان انقلاب، خیابان کارگر جنوبی، بین روانمهر و وحید نظری، بن بست

گشتاسب، پلاک ۸ تلفن: ۶۶۴۸۴۷۱۵ (۰۲۱)

تهران، خیابان انقلاب، خیابان منیری جاوید (اردیبهشت) نبش خیابان نظری، شماره ۱۴۲

تلفکس: ۶۶۴۰۰۲۲۰ - ۶۶۴۰۰۱۴۴ (۰۲۱)

<http://press.um.ac.ir>

Email: press@um.ac.ir

فهرست مطالب

۹	مقدمه مترجم
۱۰	مقدمه مؤلف
۱۵	فصل ۱. آواشناسی رایانه‌ای
۱۵	چکیده
۱۵	۱-۱ مقدمه
۱۸	۲-۱ آواها
۱۹	۱-۲-۱ آواهای گفتاری
۲۲	۲-۲-۱ آواهای آکوستیک
۲۶	۳-۲-۱ آواهای شنیداری
۲۷	۳-۱ برنامه‌های کاربردی و فناوری‌های مرتبط با آواها
۲۹	۱-۳-۱ تشخیص خودکار گفتار
۳۱	۲-۳-۱ تولید خودکار گفتار
۳۲	۳-۳-۱ شناخت ماشینی گوینده
۳۳	۴-۱ نتیجه‌گیری
۳۴	پانویس‌ها
۳۵	منابع عربی
۳۵	منابع غیرعربی
۳۷	فصل ۲. تحلیل صرفی
۳۷	چکیده
۴۰	۱-۲ ویژگی‌های صرف عربی [۱۲]
۴۱	۲-۲ رایانه و تقلید فکر انسان
۴۱	۳-۲ تحلیل صرفی
۴۲	۴-۲ تحلیلگر صرفی ماشینی
۴۳	۵-۲ همگامی نحو و صرف در پردازش ماشینی
۴۳	۶-۲ اهمیت تحلیل صرفی

- ۴۴..... ۷-۲ هدف از ساخت تحلیلگرهای صرفی ماشینی در زبان عربی.
- ۴۵..... ۸-۲ ارائه نتایج تحلیل.....
- ۴۵..... ۹-۲ مراحل عملکرد تحلیلگر صرفی ماشینی.....
- ۴۶..... ۱۰-۲ نگاه تاریخی به تحلیل صرفی ماشینی در زبان عربی.....
- ۵۱..... ۱۱-۲ روش‌های تحلیل صرفی ماشینی [۲۵].....
- ۵۲..... ۱۲-۲ قوانین و مقرراتی که یاریگر ساخت تحلیلگرهای صرفی [۲۶] هستند.....
- ۵۳..... ۱۳-۲ مشکلاتی که ساخت تحلیلگر صرفی دقیق کلمات و متون زبان عربی با آن‌ها مواجه است.....
- ۵۳..... ۱-۱۳-۲ مشکلات زبانی.....
- ۵۶..... ۲-۱۳-۲ مشکلات رایانه‌ای.....
- ۵۶..... ۱۴-۲ کیفیت توصیف قواعد صرفی برای ساخت تحلیلگر صرفی ماشینی.....
- ۵۷..... ۱۵-۲ ملزومات ساخت تحلیلگر صرفی ماشینی [۳۲].....
- ۵۷..... ۱۶-۲ ناتوانی تحلیلگرهای انگلیسی از درک ویژگی‌های زبان عربی.....
- ۵۸..... ۱۷-۲ چرا تحلیلگرهای صرفی جهانی بر زبان عربی پیشی گرفتند؟.....
- ۵۹..... ۱۸-۲ اصول پیشنهادی ساخت تحلیلگر صرفی دقیق زبان عربی.....
- ۶۰..... ۱۹-۲ هدف نهایی ما از ساخت تحلیلگر صرفی رایانه‌ای.....
- ۶۰..... ۲۰-۲ نتیجه‌گیری.....
- ۶۱..... پانویس‌ها.....
- ۶۳..... منابع.....

فصل ۳. تحلیل نحوی..... ۶۵

- ۶۵..... چکیده.....
- ۶۶..... ۱-۳ مقدمه.....
- ۶۶..... ۱-۱-۳ توصیف نحوی.....
- ۶۶..... ۱-۱-۳ ساختار ذهنی نحو.....
- ۶۷..... ۲-۱-۳ تحلیل نحوی در چهارچوب منطق ریاضی.....
- ۷۱..... ۲-۳ چشم‌اندازهای تحلیل نحوی رایانه‌ای.....
- ۷۳..... ۳-۳ اهمیت تحلیل نحوی رایانه‌ای.....
- ۷۴..... ۴-۳ مراحل تحلیل نحوی رایانه‌ای.....
- ۷۴..... ۱-۴-۳ متن خام/پیکره زبانی.....
- ۷۵..... ۲-۴-۳ تجزیه متون.....
- ۷۶..... ۱-۲-۴-۳ تجزیه در سطح جمله.....
- ۷۶..... ۲-۲-۴-۳ تجزیه در سطح واحدها/عناصر اصلی.....

۷۷.....	۳-۲-۴-۳ تجزیه در سطح واحدها/عناصر فرعی.....
۷۸.....	۳-۴-۳ برچسب گذاری اجزای کلام.....
۷۹.....	۴-۴-۳ حاشیه نویسی نحوی.....
۷۹.....	۳-۴-۴-۱ بازنمایی نحوی.....
۸۳.....	۳-۴-۴-۲ محتوای نحوی.....
۸۳.....	۳-۴-۴-۳ نظریه نحوی.....
۸۵.....	پانویس ها.....
۸۷.....	منابع عربی.....
۸۸.....	منابع غیرعربی.....

فصل ۴. تحلیل معنایی..... ۹۱

۹۱.....	چکیده.....
۹۲.....	۴-۱ مقدمه.....
۹۲.....	۴-۲ تعریف.....
۹۲.....	۴-۳ تحلیل معنایی در زبان شناسی رایانشی.....
۹۴.....	۴-۳-۱ معنای حرفی یا فعلی.....
۹۴.....	۴-۳-۲ اصطلاحات.....
۹۵.....	۴-۴ ابهام زدایی معنایی.....
۹۶.....	۴-۴-۱ موارد زبانی لازم در سیستم های ابهام زدایی معنایی.....
۹۷.....	۴-۴-۲ ابهام زدایی معنایی در زبان عربی.....
۹۸.....	۴-۵ تحلیل احساس.....
۱۰۰.....	۴-۵-۱ امتیازات و چالش های تحلیل احساس و شبکه های اجتماعی.....
۱۰۲.....	۴-۶ هستان شناسی.....
۱۰۳.....	۴-۷ تلاش های بارز در تحلیل معنایی زبان عربی.....
۱۰۴.....	پانویس ها.....
۱۰۵.....	منابع عربی.....
۱۰۵.....	منابع غیرعربی.....

فصل ۵. تجزیه و تحلیل متون..... ۱۰۷

۱۰۷.....	چکیده.....
۱۰۸.....	۵-۱ تعریف.....
۱۰۹.....	۵-۲ نقش کلان داده ها.....

۱۱۰ ۳-۵ سطوح تجزیه و تحلیل متن
۱۱۱ ۴-۵ مراحل تجزیه و تحلیل متون
۱۱۱ ۱-۴-۵ انتخاب مطالعه موردی
۱۱۲ ۲-۴-۵ مشخص کردن سؤال پژوهش یا پروژه
۱۱۲ ۳-۴-۵ انتخاب و جمع‌آوری اسناد و نمونه‌های متنی
۱۱۴ ۴-۴-۵ تکنیک استدلال منطقی
۱۱۴ ۵-۵ منابع واژگانی
۱۱۶ ۶-۵ پردازش رایانه‌ای متون
۱۱۶ ۱-۶-۵ تقسیم‌بندی متن به کلمات
۱۱۷ ۲-۶-۵ ساقه‌یابی کلمه
۱۱۸ ۳-۶-۵ آمارگیری متون
۱۱۸ ۴-۶-۵ برچسب‌گذاری اجزای کلمات
۱۱۹ ۵-۶-۵ برچسب‌گذاری اسامی موجودیت‌ها
۱۲۰ ۶-۶-۵ مدل‌های زبانی
۱۲۰ ۷-۶-۵ برنامه‌های پردازش رایانه‌ای متون
۱۲۱ ۷-۵ برنامه‌های کاربردی تجزیه و تحلیل متون
۱۲۱ ۱-۷-۵ طبقه‌بندی متون
۱۲۲ ۲-۷-۵ استخراج اطلاعات
۱۲۳ ۱-۲-۷-۵ استخراج نام موجودیت
۱۲۴ ۲-۲-۷-۵ استخراج روابط
۱۲۵ ۳-۷-۵ بازیابی اطلاعات
۱۲۷ ۴-۷-۵ تجزیه و تحلیل احساس
۱۲۸ ۱-۴-۷-۵ منابع داده‌های فرهنگ لغت برای تجزیه و تحلیل احساس
۱۳۱ ۲-۴-۷-۵ سیستم‌های تحلیل احساس
۱۳۲ ۸-۵ نتیجه‌گیری
۱۳۲ پانویس‌ها
۱۳۳ منابع
۱۳۵ فصل ۶. غلط‌یاب املائی
۱۳۵ چکیده
۱۳۶ ۱-۶ مقدمه
۱۳۷ ۲-۶ غلط‌یاب املائی به زبان عربی

فهرست مطالب ۷

۱۳۷.....	۱-۲-۶ زبان عربی و مشکلات قواعد املا.....
۱۳۹.....	۲-۲-۶ اشتباهات املائی رایج.....
۱۴۲.....	۳-۲-۶ غلط‌یاب املائی.....
۱۴۳.....	۳-۶ تشخیص خطا.....
۱۴۷.....	۴-۶ اصلاح خطاها.....
۱۴۹.....	۵-۶ اشاره پایانی به ایده کلی برخی از تکنیک‌های پیشرفته و برخی از منابع.....
۱۵۰.....	تشکر و قدردانی.....
۱۵۱.....	پانویس‌ها.....
۱۵۱.....	منابع.....
۱۵۳.....	اصطلاحات فارسی/انگلیسی.....
۱۶۳.....	نمایه.....

press.um.ac.ir

مقدمه مترجم

به نام خداوند لوح و قلم حقیقت نگار وجود و عدم

آنچه تقدیم نگاه عین و ذهن شما صاحبان اندیشه می‌شود، ترجمه‌ای است از کتاب مدخل إلى اللسانیات الحاسوبیة اثر دکتر منصور بن محمد الغامدی و همکاران که تاکنون ترجمه فارسی از آن برای علاقه‌مندان ارائه نشده است و این مهم برای نخستین بار ترجمه و چاپ می‌شود.

در این مجموعه سعی و تلاش بر این بوده است که ضمن ارائه ترجمه روان و قابل فهم، رعایت امانت از اصل کتاب متابعت شود. با توجه به اهمیت موضوع کتاب که در ارتباط با زبان‌شناسی رایانشی است، باید دانست در عصر حاضر نه تنها از سوی مخاطبان علوم رایانه، بلکه از سوی مخاطبان سایر رشته‌ها این شاخه از علم نیز مورد توجه فراوانی در حوزه اندیشه علوم زبان‌شناسی و رایانه قرار گرفته است و به‌عنوان یک تخصص میان‌رشته‌ای با بهره‌گیری از روش‌های آماری و قاعده‌بنیاد به مدل‌سازی زبان طبیعی می‌پردازد. همچنین، این علم با سابقه محدود خود در نقشه جامع علمی کشور نیز از جمله فناوری‌های دارای اولویت تعریف شده است. این توصیفات انگیزه مهمی شد برای شروع ترجمه این کتاب تا به این روش بتوانیم، سهمی در پاسخ به نیاز روز علاقه‌مندان این دانش داشته باشیم. ذکر این نکته حائز اهمیت است که اگرچه مطالب این کتاب برای زبان عربی ارائه شده‌اند، لکن کلیات آن برای تمامی زبان‌ها مانند زبان فارسی نیز قابل تعمیم است و البته این تخصص به صورت نظری و کاربردی قابل بهره‌برداری می‌باشد.

در پایان لازم می‌دانیم از نویسندگان محترم این کتاب و انتشارات بین‌المللی ملک عبدالله بن عبدالعزیز لخدمة اللغة العربیة به خاطر مساعدت جهت اخذ مجوز ترجمه این کتاب تشکر و قدردانی نماییم.

مترجمان

بهار ۱۴۰۲

مقدمه مؤلف [۱]

الحمد لله رب العالمين، والصلاة والسلام على أشرف المرسلين، نبينا محمد وعلى آله وصحبه أجمعين و بعد: زبان‌شناسی رایانشی^۱ یکی از تخصص‌های بینارشته‌ای^۲ است که بین دو علم مستقل قرار می‌گیرد: از یک سو، با زبان‌شناسی و از سوی دیگر با دانش رایانه در ارتباط است. مارتین کی [۲] معتقد است زبان‌شناسی رایانشی با احتیاط و تردید پا به عرصه وجود نهاده است. آغاز آن به سال ۱۹۴۹ میلادی برمی‌گردد؛ زمانی که وارن ویفر^۳ یادداشت معروف خود را نوشت و در آن به امکان ایجاد سامانه‌ای برای ترجمه ماشینی اشاره کرد. پس از آن، اولین اجلاس ترجمه ماشینی در مؤسسه فناوری ماساچوست (MIT) در سال ۱۹۵۲ میلادی برگزار شد. سپس مجله‌ای علمی با عنوان ترجمه ماشینی در سال ۱۹۵۴ میلادی منتشر شد، اما کاربرد اصطلاح زبان‌شناسی رایانشی به اواسط دهه شصت برمی‌گردد و به احتمال زیاد دیوید هیز^۴ زمانی که عضو کمیته مشورتی ماشینی‌سازی زبان در آکادمی ملی علوم در ایالات متحده آمریکا بود، نخستین کسی است که از نام زبان‌شناسی رایانشی استفاده کرد. پس از وی نیز تعدادی دیگر از متخصصان با ورود به این عرصه، در بروز و ظهور آن نقش به‌سزایی داشتند. از جمله می‌توان به نوآم چامسکی^۵ و جان کوک^۶ اشاره کرد. باین حال، امروزه این تخصص از اهمیت روزافزونی برخوردار است؛ زیرا نقش قابل توجهی در پیشرفت حاصل از پردازش زبان طبیعی^۷ و هوش مصنوعی^۸ دارد و همچنان یکی از پایه‌های اساسی ساخت برنامه‌های کاربردی رایانه‌ای که در بسیاری از جنبه‌های زندگی روزمره به کار می‌رود، می‌باشد.

نوگوس^۹ زبان‌شناسی رایانشی را شاخه‌ای از دو علم رایانه و زبان معرفی می‌کند که هدف از آن طراحی نمونه‌های ریاضی برای ترکیب‌های زبانی است تا بتواند به صورت ماشینی و از طریق رایانه، زبان را بررسی کند. وی همچنین معتقد است این علم به لحاظ زبان‌شناسی، قادر به شکل‌دهی به نظریات و نمونه‌های زبانی و یا اجرای آن‌ها با رایانه است. بنابراین از دیدگاه وی، این علم وسیله‌ای برای پیشرفت و توسعه نظریات زبان‌شناسانه جدید به کمک رایانه است [۳].

پژوهش در زمینه زبان‌شناسی رایانشی طی چند سال گذشته پیشرفت روزافزونی داشته است و نتایج آن

1. Computational Linguistics
2. Interdisciplinary
3. Warren Weaver
4. David Hays
5. Noam Chomsky

6. John Cocke
7. Natural Language Processing
8. Intelligence Artificial
9. Nugues

پژوهش‌ها به طور مستقیم به پیدایش برنامه‌های کاربردی کمک کرده تا در زمینه‌های مختلف بهره‌برداری شود. از مهم‌ترین این برنامه‌ها قابلیت مکالمه با ابزار است که به برنامه‌های دستیار شخصی هوشمند معروف است. از این میان می‌توان به برنامه‌های سیری^۱ از شرکت اپل^۲ و گوگل نو^۳ از شرکت گوگل، کورتانا^۴ از شرکت مایکروسافت^۵، آمازون ایکو^۶ از شرکت آمازون^۷ و ده‌ها سیستم مشابه اشاره کرد که تعدادی از سطوح پردازش زبانی را در یک برنامه کاربردی جمع کرده است. در این جا است که اهمیت زبان‌شناسی رایانشی در زبان عربی به‌منظور شناساندن این تخصص و شاخه‌ها و برنامه‌های کاربردی آن به خواننده، بیش از پیش آشکار می‌شود تا زمینه‌ای برای نوشتن منابع تخصصی‌ای باشد که به شاخه‌های آن به‌طور گسترده‌تر بپردازد. این کتاب بیش از همه برای دانشجویان مقطع تحصیلات تکمیلی دانشگاه‌ها و یا کسانی که به این حیطه علاقه‌مندند، اما تخصصی ندارند، مفید است؛ آن هم به این دلیل که تعدادی از شاخه‌های زبان‌شناسی رایانشی را به‌شرح زیر معرفی می‌کند:

- آواشناسی رایانشی^۸

- تحلیل صرفی^۹

- تحلیل نحوی^{۱۰}

- تحلیل معنایی^{۱۱}

- تحلیل متون^{۱۲}

- غلط‌یابی املائی^{۱۳}

نویسندگان این کتاب با تأکید بر مؤلفه تدریجی آموزش این طرح بیشتر به شرح اصطلاحات در حد امکان پرداخته‌اند. همچنین به منظور ساده‌سازی درک و تطبیق مطالب برای خواننده از مثال‌های روشن و واضح استفاده نموده‌اند. آنچه در ادامه بیان می‌شود، مختصری از محتویات کتاب با تکیه بر خلاصه‌هایی است که در آغاز هر فصل ذکر خواهد شد.

در فصل اول، منصور غامدی درمورد آواشناسی رایانشی سخن می‌گوید. ایشان در ابتدا مقدمه‌ای کلی راجع به آواشناسی ارائه می‌دهد. سپس به شاخه‌های سه‌گانه این علم، یعنی آواشناسی گفتاری، آکوستیک و شنیداری می‌پردازد. در ذیل هر شاخه مقدمه‌ای می‌آورد، آنگاه فناوری مرتبط با آن شاخه را به جهت مطالعه، پژوهش و تحلیل ذکر می‌کند. از آنجایی که این علم اساس و پایه تعدادی دیگر از علوم است، این

1. Siri
2. Apple
3. google now
4. Cortana
5. Microsoft
6. Amazon Echo
7. Amazon

8. Computational Phonetics
9. Morphological Analysis
10. Syntactic Parsing
11. Semantic Analysis
12. Texts Analytics
13. Spelling Checker

فصل برنامه‌های کاربردی فناوری‌های رایانه‌ای را برای علم آواشناسی، به‌ویژه تشخیص خودکار گفتار، تولید خودکار آن و همچنین شناخت گوینده همراه با ملزوماتی که اساس این فناوری‌هاست، ذکر می‌کند.

در فصل دوم، عبدالعزیز المهبوبی در مورد تحلیل صرفی سخن می‌گوید. وی سرآغازی کوتاه از ویژگی‌های صرف عربی ارائه می‌دهد. آنگاه به مفهوم و تعریف تحلیل صرفی ماشینی و پایگاه‌های داده‌های مرتبط با تحلیلگر صرفی می‌پردازد و از اصول و مبانی ساخت یک تحلیلگر صرفی سخن به میان می‌آورد و ضمن اشاره به تاریخچه تحلیل صرفی ماشینی و معرفی مهم‌ترین تحلیلگرهای صرفی در زبان عربی، بر اهمیت برنامه‌های کاربردی رایانه‌ای در تحلیل صرفی تأکید می‌نماید. پس از آن، به مجموعه‌ای از قوانین و مقرراتی که به ساخت تحلیلگرهای صرفی کمک می‌کنند، اشاره کرده و آن‌ها را به قوانین و مقررات شکلی و معنایی تقسیم می‌نماید.

سپس به مشکلاتی که در ساخت تحلیلگر دقیق صرفی کلمات و متون زبان عربی و یا در شیوه ارائه نتایج آن و کیفیت توصیف و تعریف قواعد صرفی برای ساخت تحلیلگر ماشینی پیش می‌آید، اشاره می‌کند و در پایان ضمن توضیح مراحل ساخت یک تحلیلگر ماشینی و ملزومات آن در زبان عربی، به کاستی‌های درک ویژگی‌های زبان عربی در تحلیلگرهای انگلیسی و عوامل آن نقص می‌پردازد.

در فصل سوم، احمد روبی در مورد تحلیل نحوی سخن می‌گوید و با تکیه بر شناخت شاخه‌های تحلیل نحوی (تمثیل نحوی، نظریه نحوی و محتوای نحوی) و ابزار آن، چشم‌انداز کاملی از تحلیل نحوی ماشینی را در یک چهارچوب کاربردی ارائه می‌دهد تا به‌عنوان مقدمه‌ای کوتاه برای خواننده زبان عربی باشد که او بتواند تصویری کلی از چهارچوب تحلیل نحوی رایانه‌ای را دریافت کند و در راستای تلاش برای رسیدن به این هدف، بر ساختار فرایند نحوی - ترکیبی و نقش آن در ساخت برنامه‌های کاربردی رایانه‌ای مختلف که عملکردی شبیه انسان دارد، متمرکز شود. این فصل مشتمل بر پنج محور اصلی است: مقدمه‌ای که از طریق آن راه‌های توصیف نحوی را به‌صورت ریشه‌ای توضیح می‌دهد. در ادامه با ارائه چالش‌های مواجهه با تحلیل نحوی رایانه‌ای، به اهمیت تحلیل آن به‌صورت کلی در پژوهش‌های زبانی و یا به‌صورت ویژه در زبان‌شناسی رایانشی یا پردازش زبان طبیعی می‌پردازد. سپس مراحل عملی لازم برای ساخت هر نوع فرایند تحلیل نحوی رایانه‌ای را بیان می‌نماید که می‌توان آن‌ها را به ترتیب در عناصر زیر خلاصه کرد:

(متن خام/پیکره زبانی - تجزیه متون - عنوان‌گذاری با اجزای کلام - رمزگذاری با روابط ترکیبی).

در پایان، این فصل برخی موارد تحلیل نحوی قابل اجرا در زبان عربی به همراه برنامه‌های کاربردی آن را معرفی می‌کند.

در فصل چهارم، اشراق رفاعی از تحلیل معنایی سخن می‌گوید. این فصل علاوه بر ذکر تفاوت بین معنای

حرفی و معنای کاربردی متون، تعریف کوتاهی از تحلیل معنایی و مهم‌ترین اصطلاحات مرتبط با این زمینه مانند متضادها و مترادف‌ها را بیان می‌کند و در ادامه نیز به مهم‌ترین موارد زبانی مُجاز، مانند شبکه کلمات عربی^۱، به روش تحقیقی که در پژوهش تحلیل معنایی به کار گرفته است، اشاره می‌کند. علاوه بر این، به تعدادی از مهم‌ترین برنامه‌های کاربردی در این زمینه نیز می‌پردازد که تحلیل عواطف، ابهام‌زدایی معنایی همراه با تعریف هر کدام از آن‌ها و ارائه مهم‌ترین اقدامات پژوهشی و کاربردی انجام‌شده را در برمی‌گیرد. این فصل همچنین هستان‌شناسی را به‌عنوان یک مفهوم مهم در تحلیل معنایی بررسی می‌کند و به پژوهش در کلمات مجرد و روابط بین آن‌ها به‌لحاظ معنا اهتمام ویژه دارد و در نهایت به مهم‌ترین تلاش‌های پژوهشی در زمینه تحلیل معنایی که مرتبط با زبان عربی است و همچنین مجموعه‌های پژوهشی مهمی که پیش‌از این حتی در دورترین نقاط دنیا انجام شده است، اشاره می‌کند تا از این رهگذر خواننده با جدیدترین پژوهش‌های انجام‌شده در این حوزه آشنا شود.

در فصل پنجم، صلاح‌ناجم درمورد تحلیل متون صحبت می‌کند و به اهمیت آن به‌عنوان یک برنامه کاربردی اساسی از بین برنامه‌های کاربردی پردازش رایانه‌ای زبان طبیعی می‌پردازد. در واقع، این فصل به درک و شناخت مهمی از متون آزاد فاقد ساختار منظم^۲ کمک می‌کند. همچنین به توسعه بیش‌ازپیش در کلان داده‌ها^۳ اشاره دارد که کمیت‌های بسیار زیادی را از داده‌های متنی استخراج می‌کند؛ به‌عنوان نمونه، می‌توان به مشارکت‌ها و گفت‌وگوهای رسانه‌های اجتماعی اشاره کرد؛ زیرا تحلیل این داده‌ها، با تهیه برنامه‌ها و متن‌های تحلیلی و زبان‌های برنامه‌نویسی و ابزارها و الگوریتم‌های^۴ خاصی امکان‌پذیر است که برای تعامل با این حجم فراوان از داده‌های متنی مناسب باشند و به همین دلیل اهمیت تحلیل متون به‌عنوان یک زمینه بینارشته‌ای که با بیشتر رشته‌های دانشگاهی مانند رایانه، زبان‌شناسی رایانشی، بازیابی اطلاعات^۵، تحلیل داده‌ها^۶، یادگیری ماشینی^۷ و آمار^۸ در ارتباط است، آشکار می‌شود. همچنین، این فصل درباره اهمیت کلان داده‌ها و سطوح و مراحل تحلیل متون سخن می‌گوید. سپس به بحث در رابطه با پردازش ماشینی متون و مراحل آن می‌پردازد و در پایان برخی از برنامه‌های کاربردی همچون تقسیم‌بندی متون، استخراج اطلاعات و تحلیل احساسات را بررسی می‌کند.

در فصل ششم، ولید صانع از غلط‌یابی املائی سخن می‌گوید و مهم‌ترین چالش‌هایی که توسعه‌دهندگان غلط‌یاب‌های املائی زبان عربی با آن مواجه می‌شوند را بیان می‌کند. با توجه به این که زبان عربی از جمله زبان‌هایی است که در بسیاری از سیستم‌عامل‌ها، رایانه‌ها و نرم‌افزارها پشتیبانی می‌شود،

1. Arabic WordNet
2. Unstructured Data
3. Big Data
4. Algorithms

5. Information Retrieval
6. Data Mining
7. Machine Learning
8. Statistics

شرکت‌های بزرگ جهانی اقدام به توسعه غلط‌یاب‌های املائی زبان عربی کرده‌اند. از چالش‌هایی که توسعه‌دهندگان برنامه‌های کاربردی غلط‌یاب‌های املائی زبان عربی با آن مواجه می‌شوند مربوط به تفاوت صیغه‌های املائی از لحاظ زمانی و مکانی است.

این فصل به ابزارهای شناسایی خطاهای املائی و اشکالات آن نیز می‌پردازد و مهم‌ترین راه‌های تصحیح خطاهای املائی را بیان می‌کند. همچنین با نگاهی گذرا به برخی نظریات ارائه‌شده در پژوهش‌های توسعه غلط‌یاب‌های املائی، برخی منابع اساسی که برای خواننده مفید است را ارائه می‌دهد.

در پایان پس از حمد خداوند متعال، از مسؤولان مرکز بین‌المللی خدمت به زبان عربی ملک عبدالله بن عبدالعزیز سپاسگزاری می‌کنم که حمایت پیوسته از گروه تألیف داشتند تا این کتاب به بهترین شکل تقدیم گردد و منبعی برای تلاشگران این میدان و آغازی برای دانشجویان و به‌ویژه دانشجویان تحصیلات تکمیلی باشد. همچنین از تمامی کسانی که در تألیف فصل‌های این کتاب مشارکت داشتند و با ایده و نظریات سازنده خود و نیز تعهد و التزام کاری در مسیر تحقق اهداف این کتاب و نیز نظم دادن به موضوعات آن گام برداشتند، تشکر ویژه دارم.

عبدالله بن یحیی الفیفی

ریاض - ۱۷ ربیع الثانی ۱۴۳۸

ayjfaifi@gmail.com