

روش‌ها و چالش‌های برچسب‌گذاری
اجزای کلام و ارائه روش جدید

تألیف

فرهنگ فیروزی



انتشارات فرهوش

۱۳۹۹

سرشناسه : فیروزی، فرهنگ، ۱۳۶۷ -
 عنوان و نام پدیدآور : روش‌ها و چالش‌های برچسب‌گذاری : اجزای کلام و ارائه روش جدید/تألیف فرهنگ فیروزی.
 مشخصات نشر : تهران: فرهوش، ۱۳۹۸.
 مشخصات ظاهری : ۹۴ص.
 شابک : ۹۷۸-۶۰۰-۴۹۵-۶۹۲-۵
 وضعیت فهرست نویسی : فیبا
 موضوع : فارسی -- صرف
 موضوع : Persian language -- Morphology
 موضوع : فارسی -- نحو
 موضوع : Persian language -- Syntax
 موضوع : پردازش زبان طبیعی
 موضوع : Natural language processing (Computer science)
 رده بندی کنگره : PIR ۱۰۸۴
 رده بندی دیویی : ۳۵۴
 شماره کتابشناسی ملی : ۱۱۱۰۱



روش‌ها و چالش‌های برچسب‌گذاری اجزای کلام و ارائه روش جدید
 نویسنده: فرهنگ فیروزی

ناظر فنی: سمانه قاسمی

صفحه آرا: اکرم ملک نژاد

طراح جلد: فرهنگ فیروزی

نوبت چاپ: اول - بهار ۱۳۹۹

شمارگان: ۱۰۰۰ جلد

شابک: ۹۷۸-۶۰۰-۴۹۵-۶۹۲-۵

قیمت: ۳۰۰۰۰۰ ریال

کلیه حقوق چاپ و نشر مخصوص و محفوظ ناشر است

آدرس: میدان انقلاب - ضلع جنوب شرقی - ساختمان مترجمان - پلاک ۱۷ - طبقه ۱

انتشارات: ۰۹۱۰۷۸۹۲۲۷۷ - ۶۶۴۰۱۱۶۱ دبیرخانه: ۶۶۹۱۱۷۲۶ - ۶۶۴۰۶۳۰۹

www.farhooshpub.com

دورنگار: ۸۹۷۸۴۵۵۸

فهرست مطالب

فصل ۱: مقدمه	۹
۱-۱- مقدمه	۱۱
۱-۱-۱- پردازش زبان طبیعی	۱۲
۱-۱-۲- تجزیه نحوی	۱۴
۲-۱- بیان مسئله	۱۵
۳-۱- کاربرد برجسب گذاری اجزای کلام و هدف پروژه	۱۶
۴-۱- اختار	۱۶
فصل ۲: برجسب گذاری اجزای کلام	۱۹
۱-۲- مقدمه	۲۱
۲-۲- برجسب گذاری برای کلام	۲۱
۳-۲- انواع برجسب گذاری اجزای کلام	۲۳
۳-۲-۱- برجسب گذاری اجزای کلام مبتنی بر مدل مارکوف	۲۴
۳-۲-۲- ویربی	۲۶
۳-۲-۳- برجسب گذاری اجزای کلام مبتنی بر حافظه	۲۹
۳-۲-۴- برجسب گذاری اجزای کلام مبتنی بر احتمال پیشینه	۳۰
۴-۲- مسائل و مشکلات برجسب گذاری اجزای کلام	۳۱
۴-۲-۱- مسائل و مشکلات ناشی از ابهام زبان	۳۲
۴-۲-۲- مسائل و چالش های زبان فارسی	۳۲
۴-۲-۳- مواجهه با چند معنایی و چند نقشی بودن کلمات	۳۳
۴-۲-۴- حذف کردن کلمات و عبارات به قرینه لفظی یا معنوی	۳۳
۴-۲-۵- استفاده از افعال مرکب اصطلاحات و ضرب المثل ها و شعر	۳۳
۴-۲-۶- تعیین طبقه اسمی	۳۳
۴-۲-۷- بی ترتیب بودن زبان	۳۴
۴-۲-۸- کسره اضافه و حذف آن	۳۴
۴-۲-۹- عدم تطابق اجزای جمله	۳۴
۴-۲-۱۰- وجود ساختار جملات یکسان با معانی و نقش های متفاوت	۳۵
۴-۲-۱۱- مسائل پردازش محاسباتی در پردازش متون فارسی	۳۵

۲-۵- پیکره متی..... ۳۵

۲-۵-۱- مجموعه برجسب..... ۳۷

۲-۶- کلمات ناشناخته..... ۳۹

۲-۷- جمع بندی..... ۳۹

فصل ۳: کارهای پیشین..... ۴۱

۳-۱- مقدمه..... ۴۳

۳-۲- بررسی مشکلات پردازش زبان فارسی..... ۴۳

۳-۳- بررسی کارهای انجام شده برای برجسب گذاری زبان فارسی..... ۴۵

۳-۳-۱- مدل مارکوف..... ۴۵

۳-۳-۲- تخمین احتمال پیشینه..... ۴۸

۳-۳-۳- روش مبتنی بر حافظه..... ۵۰

۳-۳-۴- سایر کارها..... ۵۲

۳-۴- جمع بندی..... ۵۴

فصل ۴: روش های پیشنهادی و ارزیابی آن..... ۵۵

۴-۱- مقدمه..... ۵۷

۴-۲- دادگان مورد استفاده..... ۵۸

۴-۳- روش های اولیه..... ۶۰

۴-۳-۱- پیاده سازی تخمین احتمال پیشینه..... ۶۱

۴-۳-۲- پیاده سازی ترکیبی..... ۶۴

۴-۳-۳- پیاده سازی جدولی..... ۶۵

۴-۴- ابزار قبلی..... ۶۶

۴-۴-۱-۱- اشاره ای به پیاده سازی مدل مارکوف..... ۶۶

۴-۴-۲- انتقال های غیر مفید..... ۶۷

۴-۴-۳- برجسب گذاری دسته بندی شده..... ۶۹

۴-۴-۴- وزن دهی فاکتورها..... ۷۰

۴-۴-۵- پیکره معکوس..... ۷۲

۴-۴-۶- فاکتورهای جدید..... ۷۳

۷۵ ۴-۶-۱ جستجوی تپه‌نوردی

۷۶ ۴-۶-۲ الگوریتم جستجوی ژنتیک

۷۸ ۴-۵ جمع‌بندی

۸۱ فصل ۵: نتیجه‌گیری و کارهای آینده

۸۳ ۵-۱ مقدمه

۸۴ ۵-۲ نتیجه‌گیری

۸۷ ۵-۳ کارهای آینده

۹۱ مراجع

www.ketab.ir