Хранение данных

Вычислительная лингвистика использует корпусы для обучения моделей машинного перевода. Корпус — это большой набор текстов, выровненных по предложениям. Чем больше корпус, тем точнее модель. Для редких языков корпусов почти нет, поэтому исследователи собирают их вручную.

Выравнивание предложений — отдельная задача. Иногда одному предложению соответствует два или три предложения другого языка, иногда предложение пропускается. Алгоритмы выравнивания учитывают длину предложений и частотность слов.

Корпус хранится в виде файлов с разметкой. Каждое предложение сопровождается номером, языком и источником. Метаданные позволяют фильтровать корпус по жанру, эпохе и качеству.

Качество корпуса проверяют выборочно: носитель языка читает случайные предложения и отмечает ошибки. Если ошибок больше порога, корпус отправляют на доработку.

Объем корпуса измеряют в миллионах слов. Для английского языка корпусы насчитывают миллиарды слов, для большинства других языков — десятки миллионов.

Корпус хранится в виде файлов с разметкой. Каждое предложение сопровождается номером, языком и источником. Метаданные позволяют фильтровать корпус по жанру, эпохе и качеству.

Выравнивание предложений — отдельная задача. Иногда одному предложению соответствует два или три предложения другого языка, иногда предложение пропускается. Алгоритмы выравнивания учитывают длину предложений и частотность слов.

Качество корпуса проверяют выборочно: носитель языка читает случайные предложения и отмечает ошибки. Если ошибок больше порога, корпус отправляют на доработку.

Корпус — это большой набор текстов, выровненных по предложениям. Чем больше корпус, тем точнее модель. Для редких языков корпусов почти нет, поэтому исследователи собирают их вручную.

Объем корпуса измеряют в миллионах слов. Для английского языка корпусы насчитывают миллиарды слов, для большинства других языков — десятки миллионов.

Выравнивание предложений — отдельная задача. Иногда одному предложению соответствует два или три предложения другого языка, иногда предложение пропускается. Алгоритмы выравнивания учитывают длину предложений и частотность слов.

Корпус хранится в виде файлов с разметкой. Каждое предложение сопровождается номером, языком и источником. Метаданные позволяют фильтровать корпус по жанру, эпохе и качеству.

Качество корпуса проверяют выборочно: носитель языка читает случайные предложения и отмечает ошибки. Если ошибок больше порога, корпус отправляют на доработку.

Объем корпуса измеряют в миллионах слов. Для английского языка корпусы насчитывают миллиарды слов, для большинства других языков — десятки миллионов.

Корпус-уникум: выравнивание

Выравнивание предложений — отдельная задача. Иногда одному предложению соответствует два или три предложения другого языка, иногда предложение пропускается. Алгоритмы выравнивания учитывают длину предложений и частотность слов.

Корпус хранится в виде файлов с разметкой. Каждое предложение сопровождается номером, языком и источником. Метаданные позволяют фильтровать корпус по жанру, эпохе и качеству.

Качество корпуса проверяют выборочно: носитель языка читает случайные предложения и отмечает ошибки. Если ошибок больше порога, корпус отправляют на доработку.

Объем корпуса измеряют в миллионах слов. Для английского языка корпусы насчитывают миллиарды слов, для большинства других языков — десятки миллионов.

Корпус — это большой набор текстов, выровненных по предложениям. Чем больше корпус, тем точнее модель. Для редких языков корпусов почти нет, поэтому исследователи собирают их вручную.

Выравнивание предложений — отдельная задача. Иногда одному предложению соответствует два или три предложения другого языка, иногда предложение пропускается. Алгоритмы выравнивания учитывают длину предложений и частотность слов.

Корпус хранится в виде файлов с разметкой. Каждое предложение сопровождается номером, языком и источником. Метаданные позволяют фильтровать корпус по жанру, эпохе и качеству.

Качество корпуса проверяют выборочно: носитель языка читает случайные предложения и отмечает ошибки. Если ошибок больше порога, корпус отправляют на доработку.

Объем корпуса измеряют в миллионах слов. Для английского языка корпусы насчитывают миллиарды слов, для большинства других языков — десятки миллионов.

Корпус-уникум: выравнивание

Выравнивание предложений — отдельная задача. Иногда одному предложению соответствует два или три предложения другого языка, иногда предложение пропускается. Алгоритмы выравнивания учитывают длину предложений и частотность слов.

Корпус хранится в виде файлов с разметкой. Каждое предложение сопровождается номером, языком и источником. Метаданные позволяют фильтровать корпус по жанру, эпохе и качеству.

Качество корпуса проверяют выборочно: носитель языка читает случайные предложения и отмечает ошибки. Если ошибок больше порога, корпус отправляют на доработку.

Объем корпуса измеряют в миллионах слов. Для английского языка корпусы насчитывают миллиарды слов, для большинства других языков — десятки миллионов.

ⵜⵉⵎⵉⵔⵉ ⵏ ⵜⵎⵣⵓⴳⵖⵉ ⵙ ⵜⵉⵏⵏⵉⵏⵏⵉ ⴷ ⵜⵉⵎⵣⵓⴳⵖⵉ ⵏ

𐤀𐤋𐤉𐤌𐤀𐤕 𐤔𐤀𐤋𐤀𐤌 𐤀𐤕 𐤔𐤉𐤓𐤕𐤔 𐤔𐤀𐤋𐤀𐤌𐤕𐤌 𐤋𐤀 𐤔𐤀𐤋𐤀𐤌 𐤔𐤀𐤋𐤀𐤌 𐤀𐤕 𐤔𐤉𐤓𐤕𐤔 𐤔𐤀𐤋𐤀

𐤊𐤓𐤔𐤌 𐤕𐤌𐤉𐤋 𐤔𐤋𐤌𐤄

Ирландский фольклор – Легенды Ирландии – Лесной Король (Король) Ирландии в лесах, Дикая Охота, Ирландский Лесной Король - Дикая Охота Ирландии, Ирландский Лесной Король, Ирландский Лесной Король (Король) Ирландии в лесах. Ирландский Король Ирландии в XVI. Ирландский Король Ирландии в XVI. Ирландский Король Ирландии в XVI. Ирландский Король Ирландии в XVI. Ирландский Король Ирландии в XVI. Ирландский Король Ирландии в XIX. Ирландский Король Ирландии в XIX. Ирландский Король Ирландии в XIX

Ритмы Ночного Города

Ночь в Ритмополисе — это не отсутствие звука, а его другая плотность. Когда солнце уходит за горизонт, город не замолкает, но меняет партитуру. Дневные ритмы — деловые, быстрые, рассчитанные на толпу — уступают место медленным, тяжелым, почти сонным ритмам, которые звучат для немногих. Ночной Ритмополис — это город, который дышит медленнее, город, который шепчет вместо того, чтобы кричать, город, который позволяет своим жителям быть собой, а не частью чьего-то расписания.

Ночной Ритмополис — это город, который дышит медленнее, город, который шепчет вместо того, чтобы кричать, город, который позволяет своим жителям быть собой, а не частью чьего-то расписания.

𐤁𐤔𐤋𐤌𐤄 𐤀𐤓𐤆 𐤄𐤀𐤓𐤆

ЖАЖМЖІАЛИ ОЗИНО ЧЖЕЖ ЗУ (ЖЕЖ ОЛИНО ЖАСІЖМЖЖЖ ⊗ЖИНЕЛСЕ ВАЕЕЕ, ИЕЖ ЗУ (ЖЕЖ ОЛИНО ЖАСІЖМЖЖЖ ВЕЕ І. ЧЛЕЛЕ Ч УЛОСУ (ЖЕЖИЖОЧЕЖ ЕЖЛЖЖЛЖЖЖ ИЕЛЖЛЖЖ ЗИЕ ЗУ ЛЕЛЕЛІЛЕЛ ОЗИНО ЖЕЖ ЗУ ЛЕЛЕЛІЛЕЛ ЗУ ЖЛЕЛЕЛЕ

ᚠᚢᚦᚨᚱᚲᚷᚹᚺᚾᛁᛃᛇᛈᛉᛊᛏᛒᛖᛗᛚᛜᛞᛟ

Ни Вы́й-Хвост хвостовидный, Ни Ни́з-Ни́з, Выжимающий Ни́з-Ни́з, Ни́з-Ни́з, Ни́з-Ни́з,