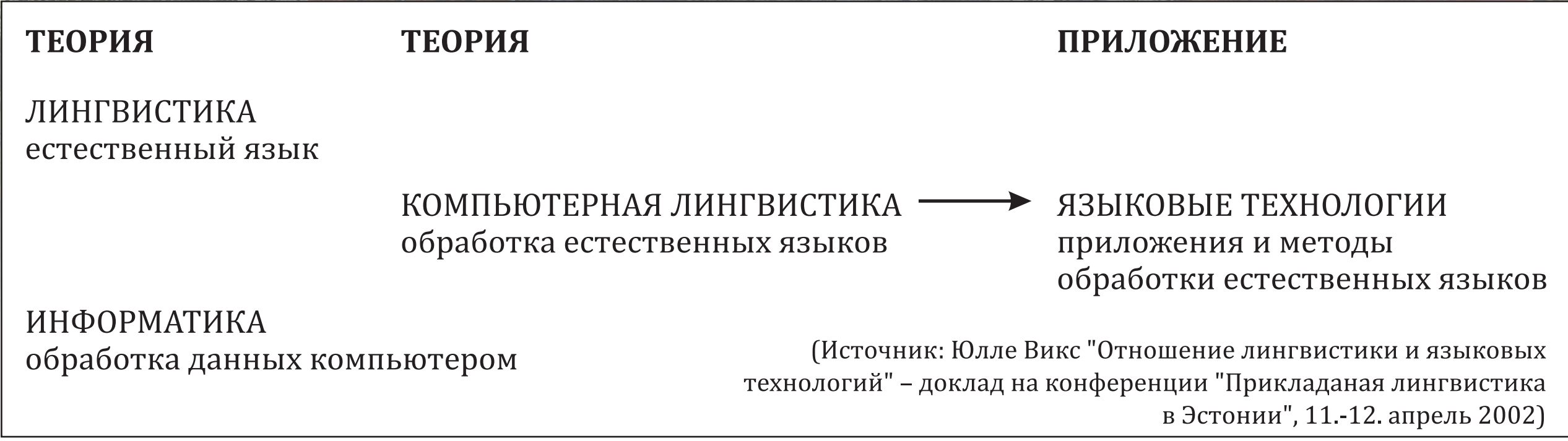


Языковые технологии в Эстонии

Что такое языковые технологии и компьютерная лингвистика?

Языковые технологии – это ветвь информатики, которая занимается обработкой естественных языков, тогда как под **компьютерной лингвистикой** подразумевается научная область-гибрид лингвистики и информатики. И та и другая область занимаются автоматической обработкой естественных языков: у компьютерной лингвистики подход более теоретический, у языковых технологий – более практический.



Халдур Ёйм



Маре Койт



Юлле Викс

Где в Эстонии сосредоточено изучение языковых технологий и компьютерной лингвистики?

В Тартуском университете существует две исследовательских группы, занимающихся копьютерной лингвистикой и языковыми технологиями. В Институте Эстонского языка занимаются синтезом речи и компьютерной лексикографией. В Таллиннском техническом университете существует лаборатория фонетики и речевых технологий, как часть института кибернетики.

Изучать эти области можно в Тартуском университете, где существуют соответствующие специальности: специальность компьютерной лингвистики при факультете философии, в рамках учебной программы эстонской и финно-угорской лингвистики, и специальность языковых технологий при факультете математики-информатики, в рамках учебной программы инфотехнологий. Также в Таллиннском техническом университете читаются курсы по речевым технологиям.

Чем занимаются компьютерная лингвистика и языковые технологии в Эстонии? Некоторые примеры:

Любой человек, занимающийся подготовкой текста на компьютере сталкивается со спеллером, или автоматическим корректировщиком правописания. Он основан на автоматичеком морфологическом анализе – на программе, которая определяет начальную форму и склонение или спряжение любого слова в тексте. Спеллер подчёркивает красной чертой как раз те слова, которые не удалось морфологически проанализировать. Также морфологический анализ (как и обратный процесс – морфологический синтез) используются для поиска информации: например, чтобы на запрос *лето* также подходил документ содержащий фразу конец лета.

Следущая ступень после корректировщика правописания – это грамматический корректировщик, который находил бы, например, пунктуационные ошибки в тексте. Для создания такой программы необходим автоматический анализ синтаксиса, способный находить части предложений и их синтаксическую структуру.

Составители словарей используют интернетовскую рабочую среду, интегрирующую в себе языковые ресурсы и приложения языковых технологий чтобы ускорить составление и редактирование словарей.

Очень ценными в будущем обещают быть диалоговые системы, т.е. пользовательские интерфейсы, поддерживающие естественные языки. Для создания диалоговой системы, отвечающей на телефонные звонки кроме распознавания речи (т.е. программы, превращающей речь в письменный текст) необходимы программы для перевода вопроса пользователя в запрос в базу данных и для представления найденной информации в естественном виде. Для этого необходимо знать, как происходит общение по телефону между людьми – как задают вопросы и отвечают на них. В конце также необходим и синтез речи для того чтобы проговорить ответ звонящему.

Здесь приведена статья газеты Postimees "Большая часть мусора, собранного в течение кампании, будет заново использована", которая будет направлена на обработку автоматического составителя резюме. Пользователь может выбрать объём резюме: 10,

Вывод автоматического составителя резюме: 30% статьи "Большая часть мусора, собранного в течение кампании, будет заново использована".

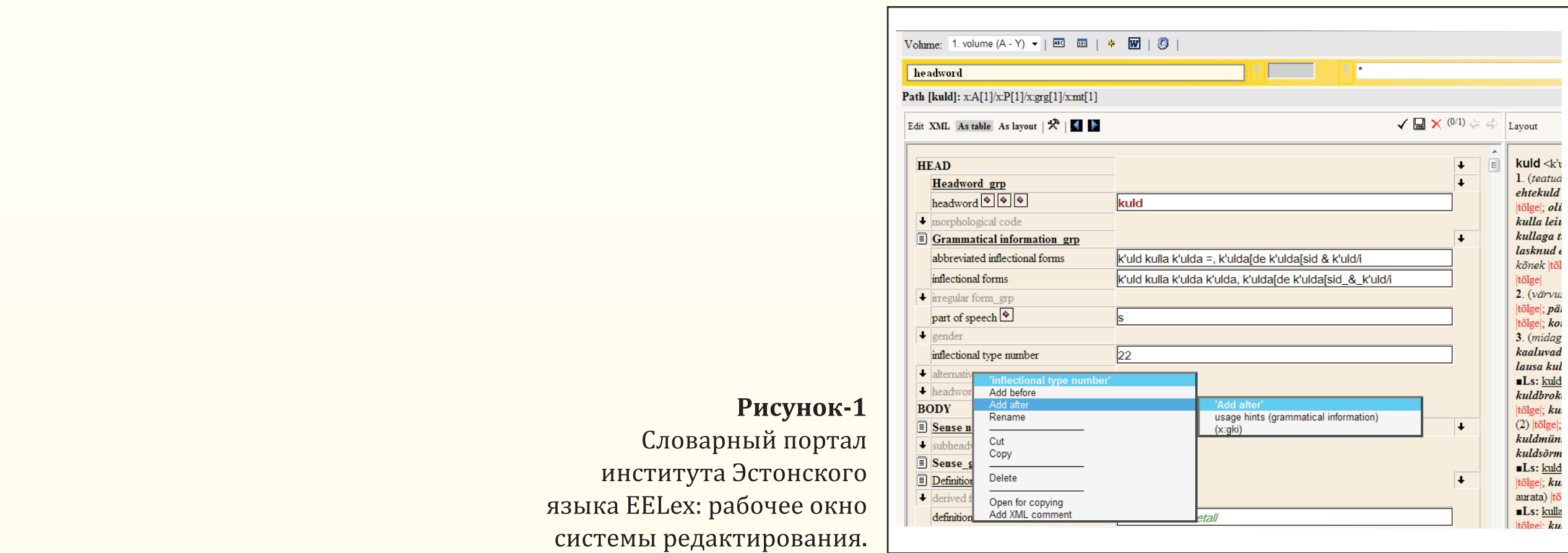
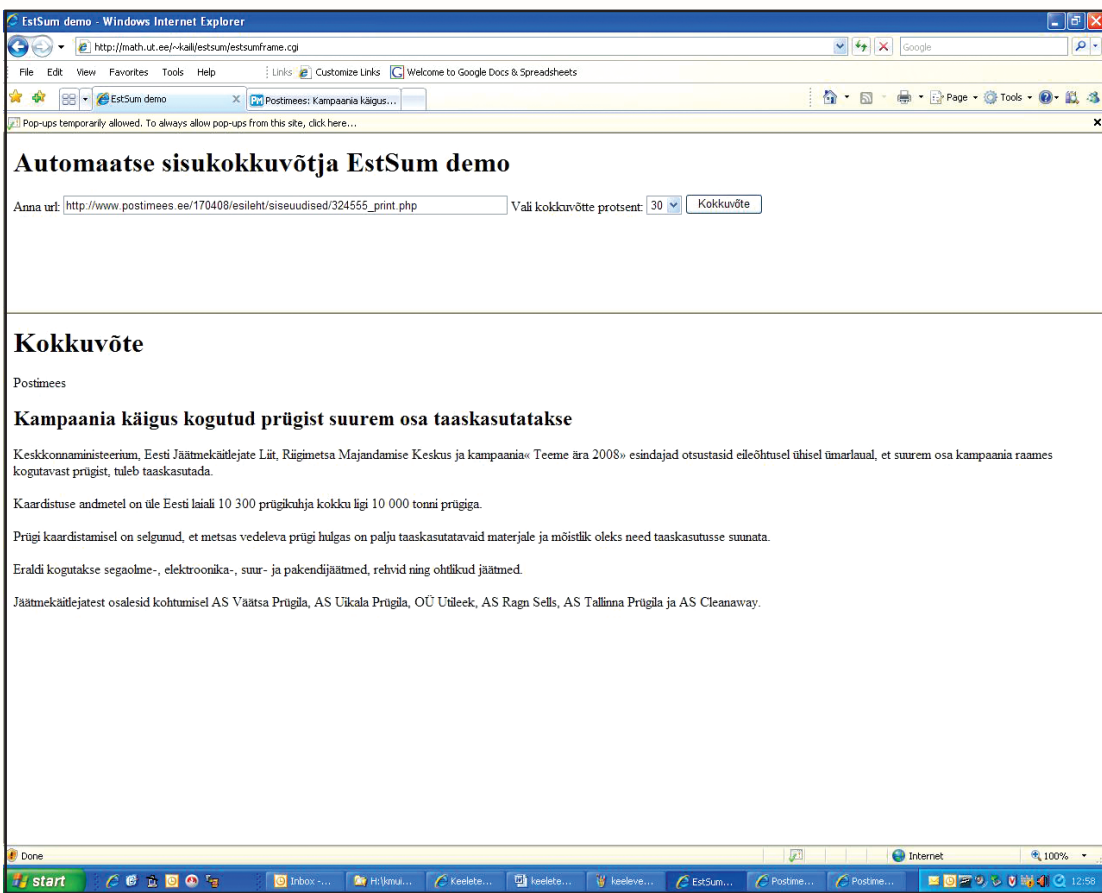
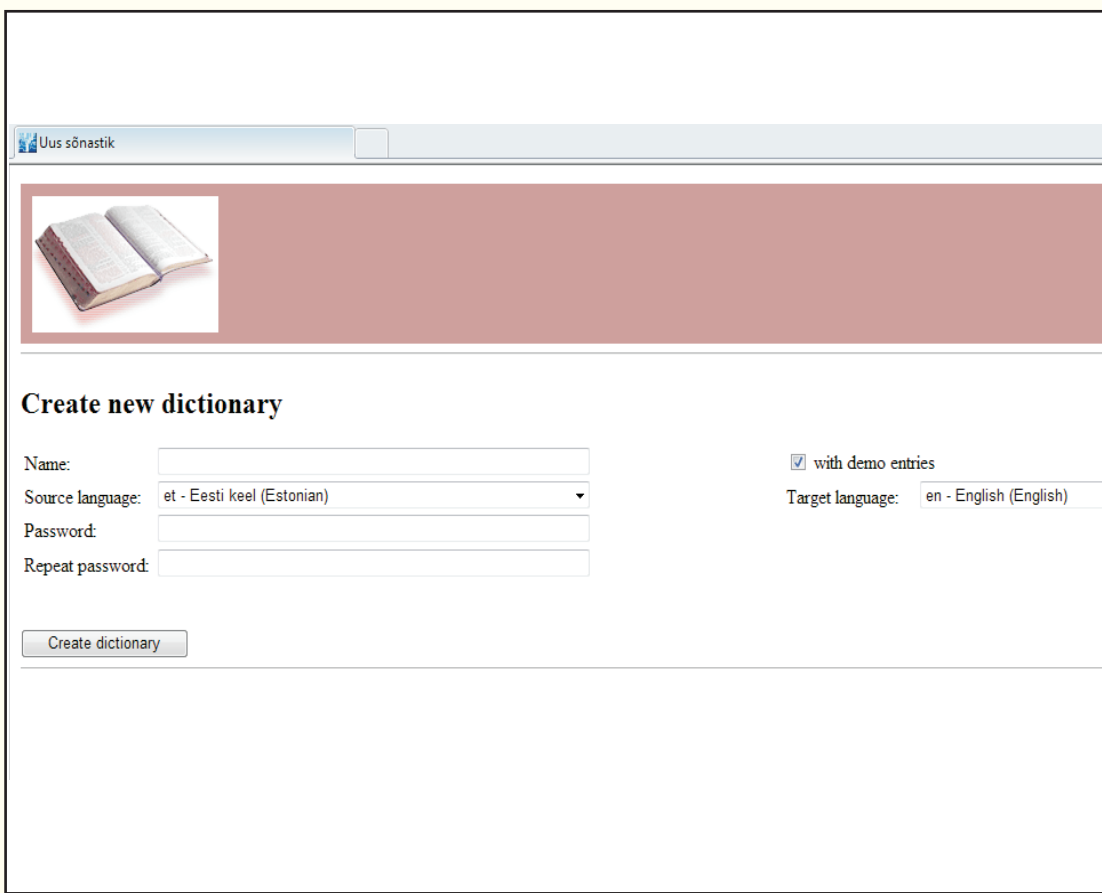


Рисунок-1
Словарный портал
института Эстонского
языка EELEX: рабочее окно
системы редактирования.

Диалоговая система "Театральный агент" даёт информацию о программах Эстонских театров.

Рисунок-2
Словарный портал
института Эстонского
языка EELEX: окно
создания нового
словаря.
loomise aken.



Интернетовские адреса:

Исследовательская группа компьютерной лингвистики: <http://www.cl.ut.ee>; Исследовательская группа языковых технологий: <http://www.cs.ut.ee/~koit/KT/>; Институт Эстонского языка: <http://www.eki.ee>; ТТИ, институт кибернетики, лаборатория фонетики и речевых технологий: <http://www.ioc.ee/>; Диалоговые системы: <http://www.dialoogid.ee> Синтез речи: <http://www.eki.ee/keeletehnoloogia/projektid/syntees/>; <http://www.phon.ioc.ee/->projektid->tekst-kõne-süntees-pimedatele> Общие порталы электронных словарей: <http://www.keeleveeb.ee>; <http://www.keelevara.ee> Словарный портал института Эстонского языка EELEX: <http://exsa.eki.ee/> Автоматический составитель резюме: <http://math.ut.ee/~kaili/estsum/estsumframe.cgi>

Изготовитель: Общество родного языка, Эстонский музей литературы, Эстонское министерство образования и науки. Текст: Кадри Муишнек. Оформление: Маргус Нымм, Центр мультимедии Тартуского университета, 2008

