

TabbyXL: система трансформации данных из произвольных электронных таблиц в реляционную форму

ШИГАРОВ АЛЕКСЕЙ

ФГБУН Институт динамики систем и теории управления СО РАН (Иркутск), Россия
e-mail: shigarov@icc.ru

БЫЧКОВ ИГОРЬ ВЯЧЕСЛАВОВИЧ

ФГБУН Институт динамики систем и теории управления СО РАН (Иркутск), Россия
e-mail: bychkov@icc.ru

РУЖНИКОВ ГЕННАДИЙ МИХАЙЛОВИЧ

ФГБУН Институт динамики систем и теории управления СО РАН (Иркутск), Россия

ПАРАМОНОВ ВЯЧЕСЛАВ ВЛАДИМИРОВИЧ

ФГБУН Институт динамики систем и теории управления СО РАН (Иркутск), Россия
e-mail: slv@icc.ru

МИХАЙЛОВ А. А.

ФГБУН Институт динамики систем и теории управления СО РАН (Иркутск), Россия

В работе обсуждаются вопросы основанной на правилах трансформации данных из произвольных электронных таблиц в реляционную форму. Определяется новая объектная модель произвольной таблицы и предметно-ориентированный язык правил анализа и интерпретации табличной структуры. Предлагается оригинальная система извлечения и трансформации данных электронных таблиц (ТАВВУХЛ). Экспериментальная оценка показывает высокую точность (от 93,7% до 100%) и полноту (от 93,9% до 99,8%) для восстановления различных функциональных единиц данных и их отношений на известном тестовом наборе 200 произвольных таблиц одного жанра (государственных статистических отчетов).