

THE GRAPHIC NARRATIVE CORPUS (GNC): DESIGN, ANNOTATION, AND ANALYSIS FOR THE DIGITAL HUMANITIES

ALEXANDER DUNST

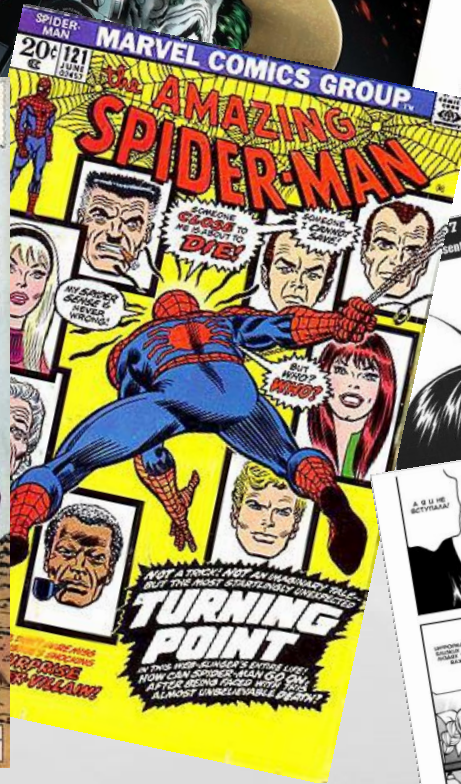
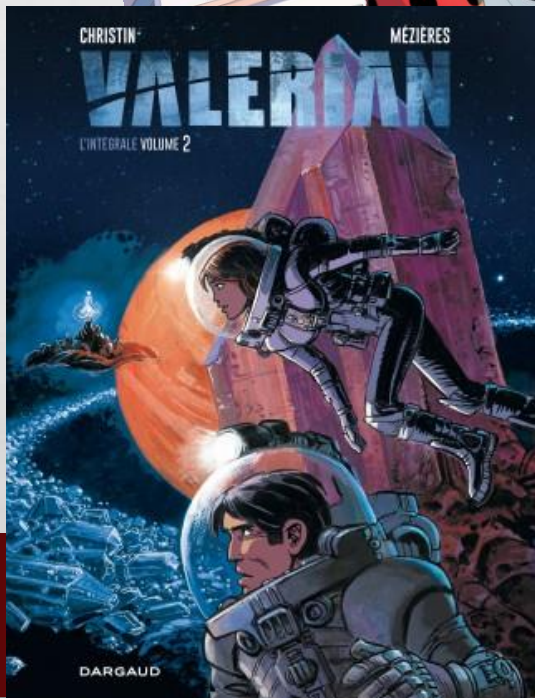
RITA HARTEL

JOCHEN LAUBROCK

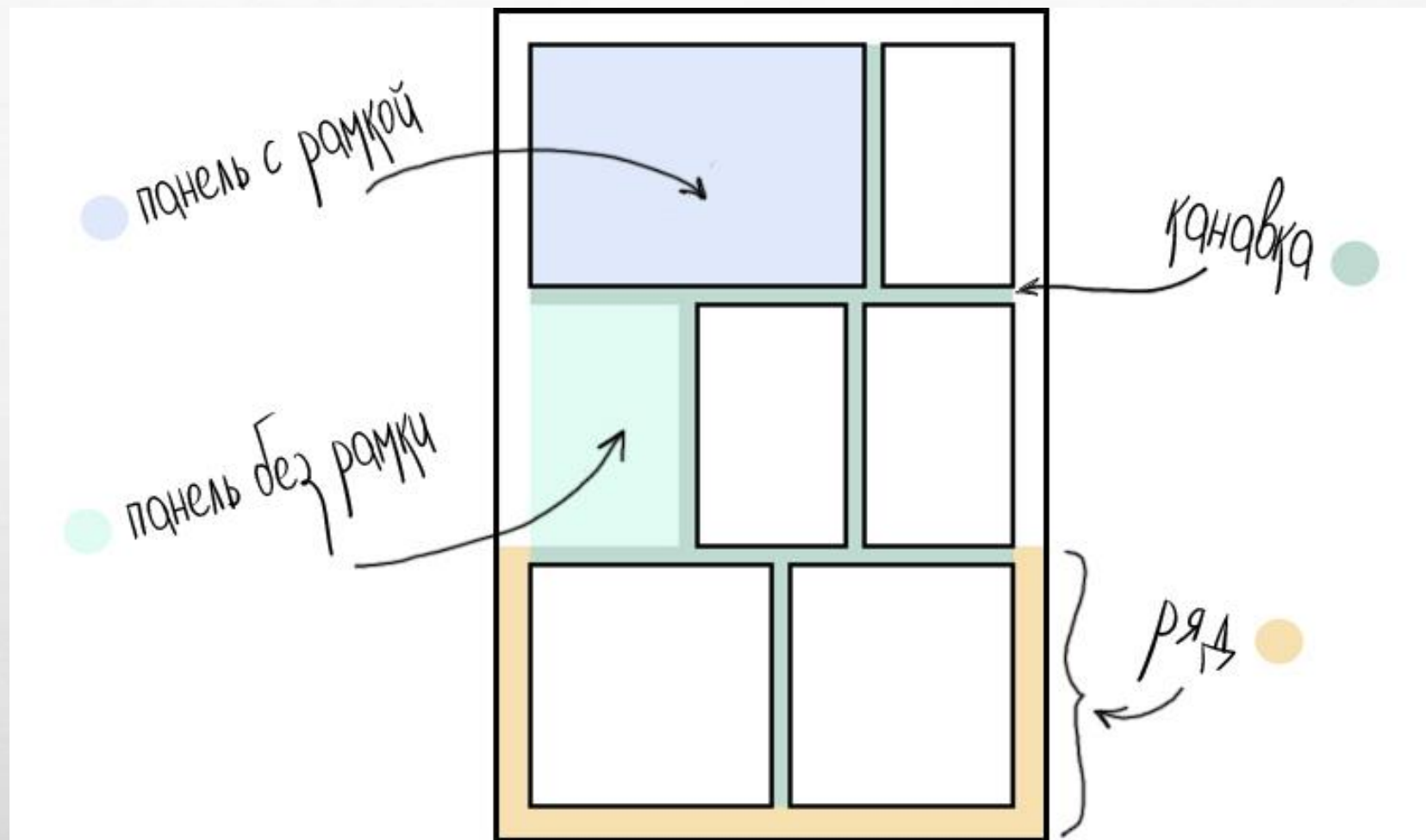
2017



Выполнила Семенова Наталья, МКЛ-171



КОМИКСЫ, ВД, МАНГА

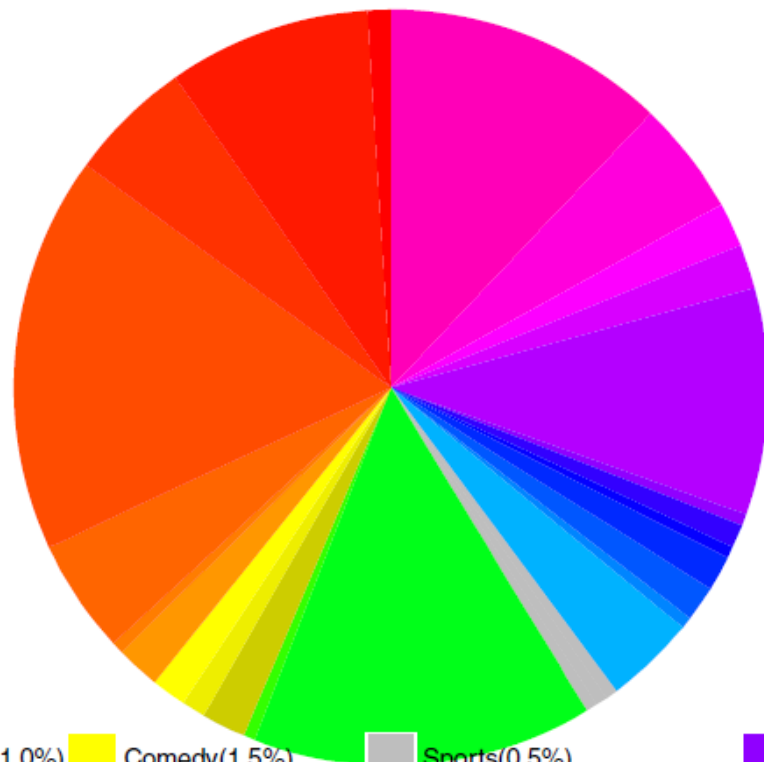


ВИЗУАЛЬНАЯ СТРУКТУРА КОМИКСА

КОРПУСЫ

- EBDTHEQUE (*BANDES DÉSSINÉES*) ONLY RUNS TO 100 PAGES
- MANGA109 CONTAINS AS MANY COMIC BOOKS WITH A TOTAL OF 21,000 PAGES
- IYYER CORPUS “CONTAINS ~1,2 MILLION PANELS DRAWN FROM ALMOST 4,000 PUBLICLY AVAILABLE COMIC BOOKS”

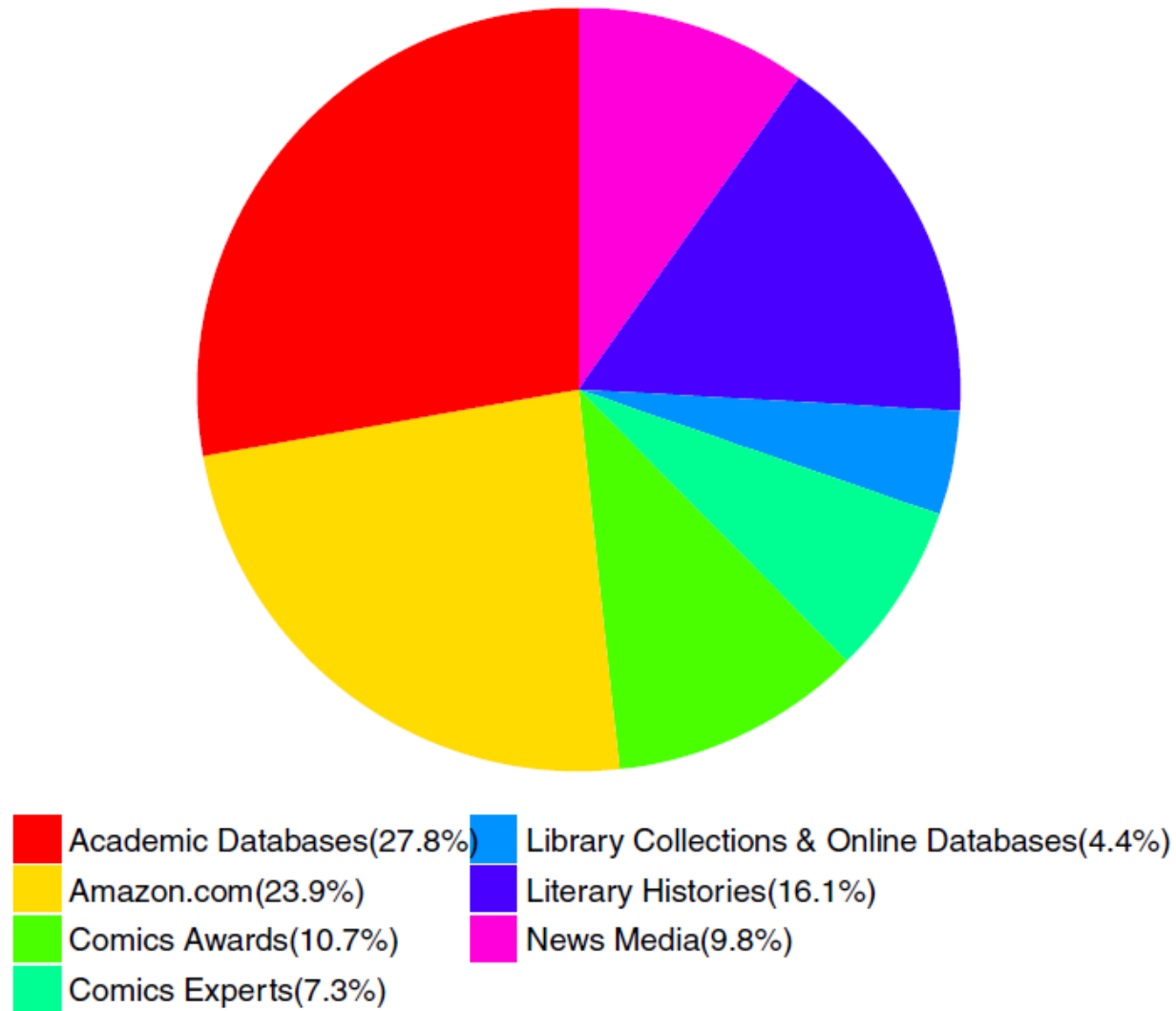
СБАЛАНСИРОВАННАЯ И РЕПРЕЗЕНТАТИВНАЯ ВЫБОРКА ТЕКСТОВ!!!



**Оцифрованы 160 полноразмерных
изданий (всего более 38,000 страниц),
обещают к 2018 дойти до 250 изданий**

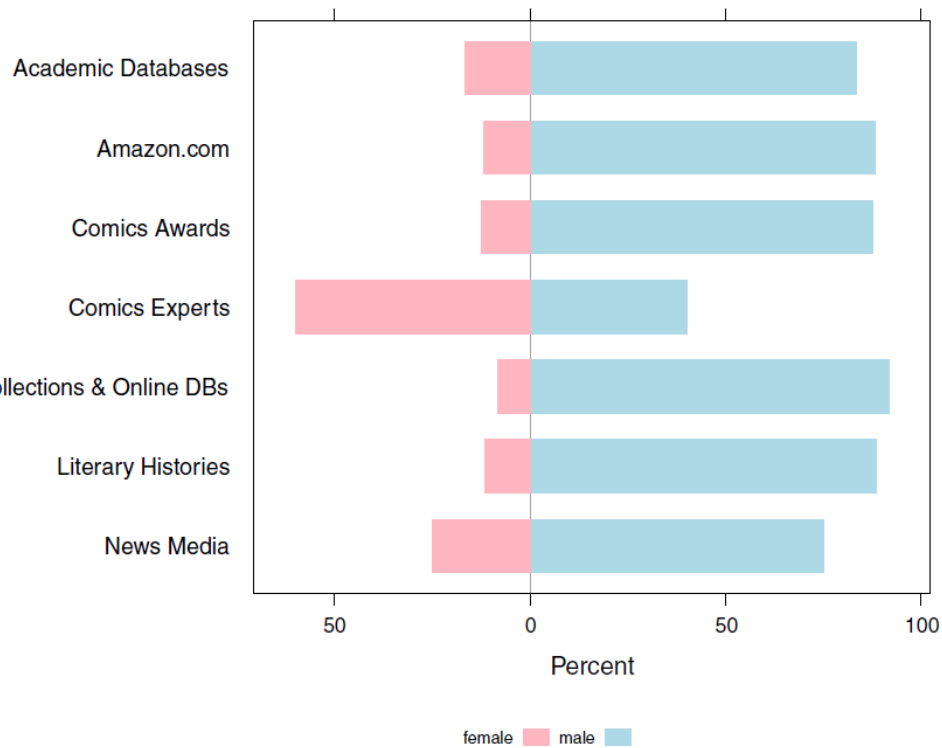
**Выбраны издания (>64 страниц)
с 1970 года**

CORPUS DESIGN

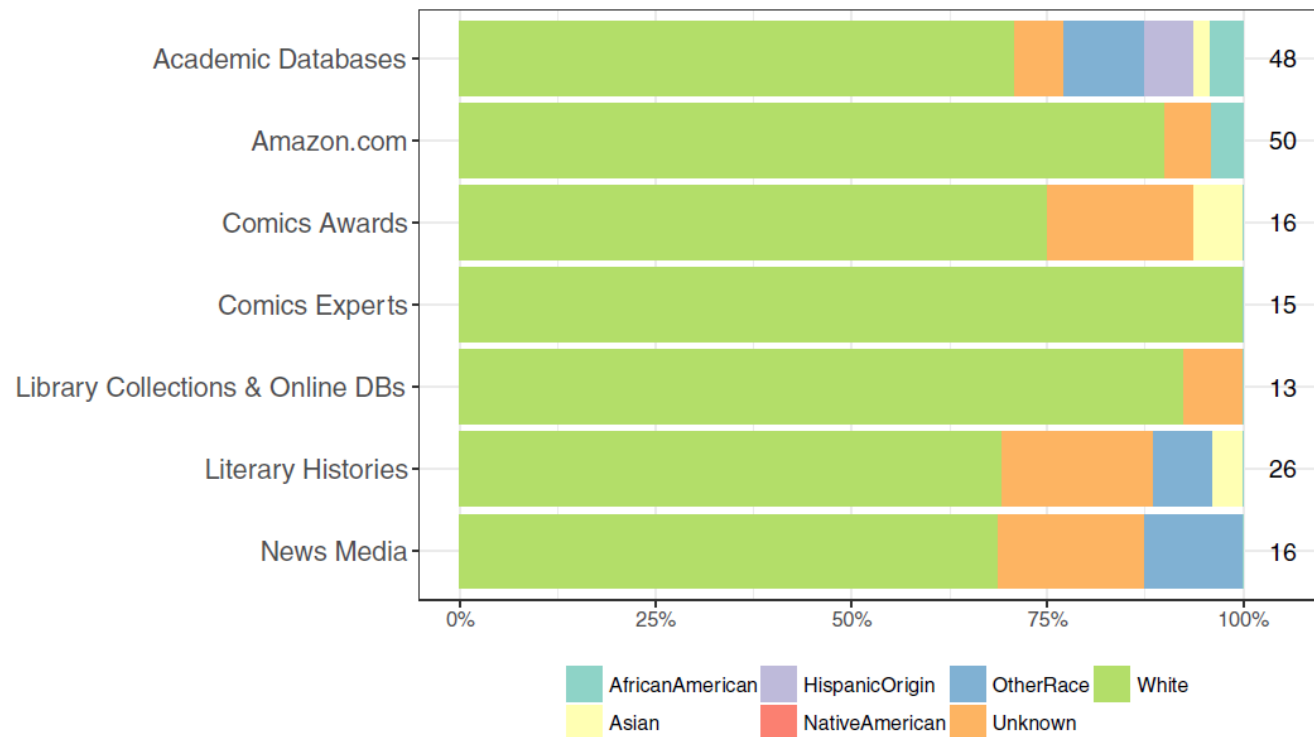


DISTRIBUTION OF TITLES PER SOURCE

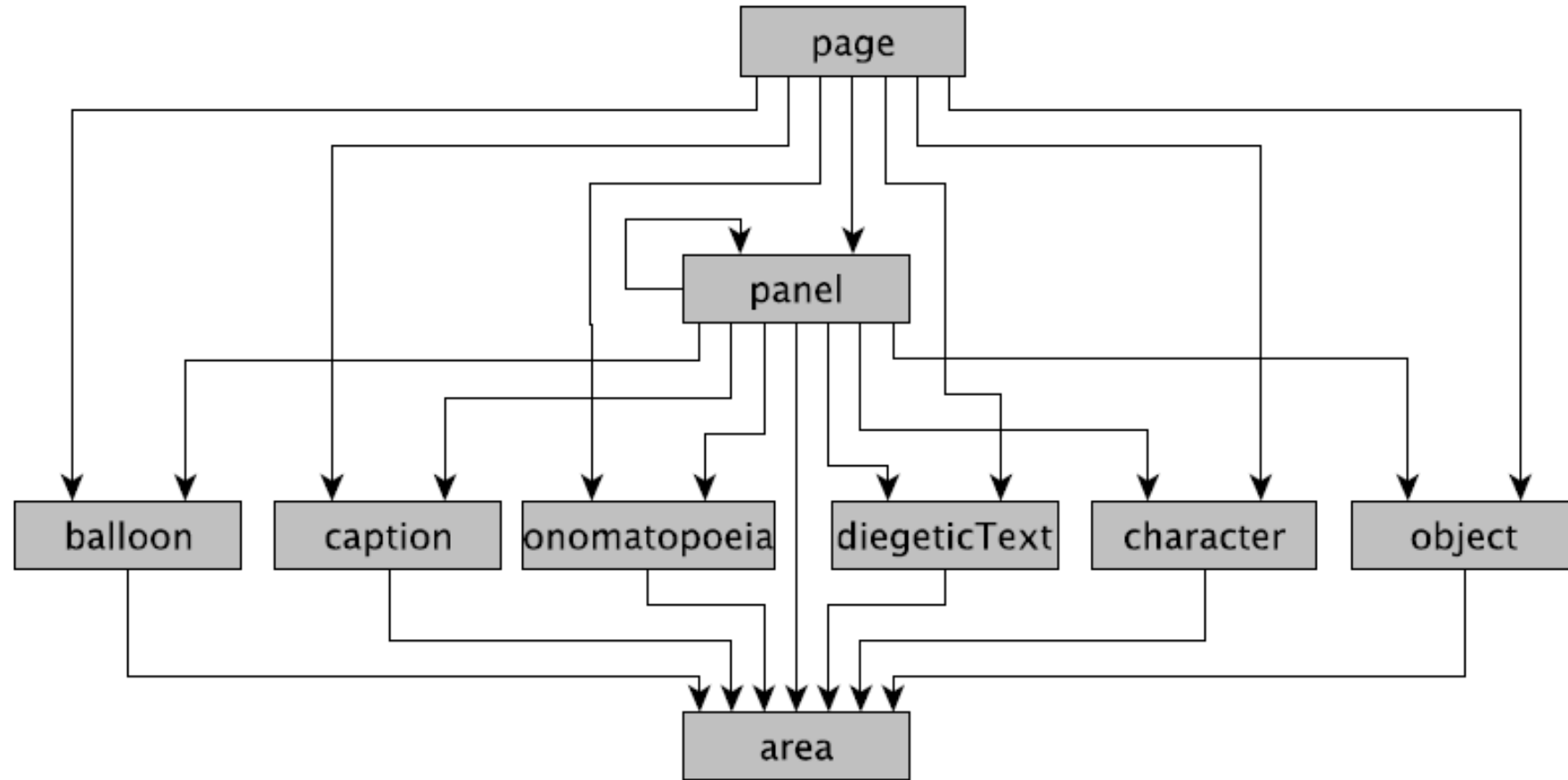
Gender per Source (US Publications)



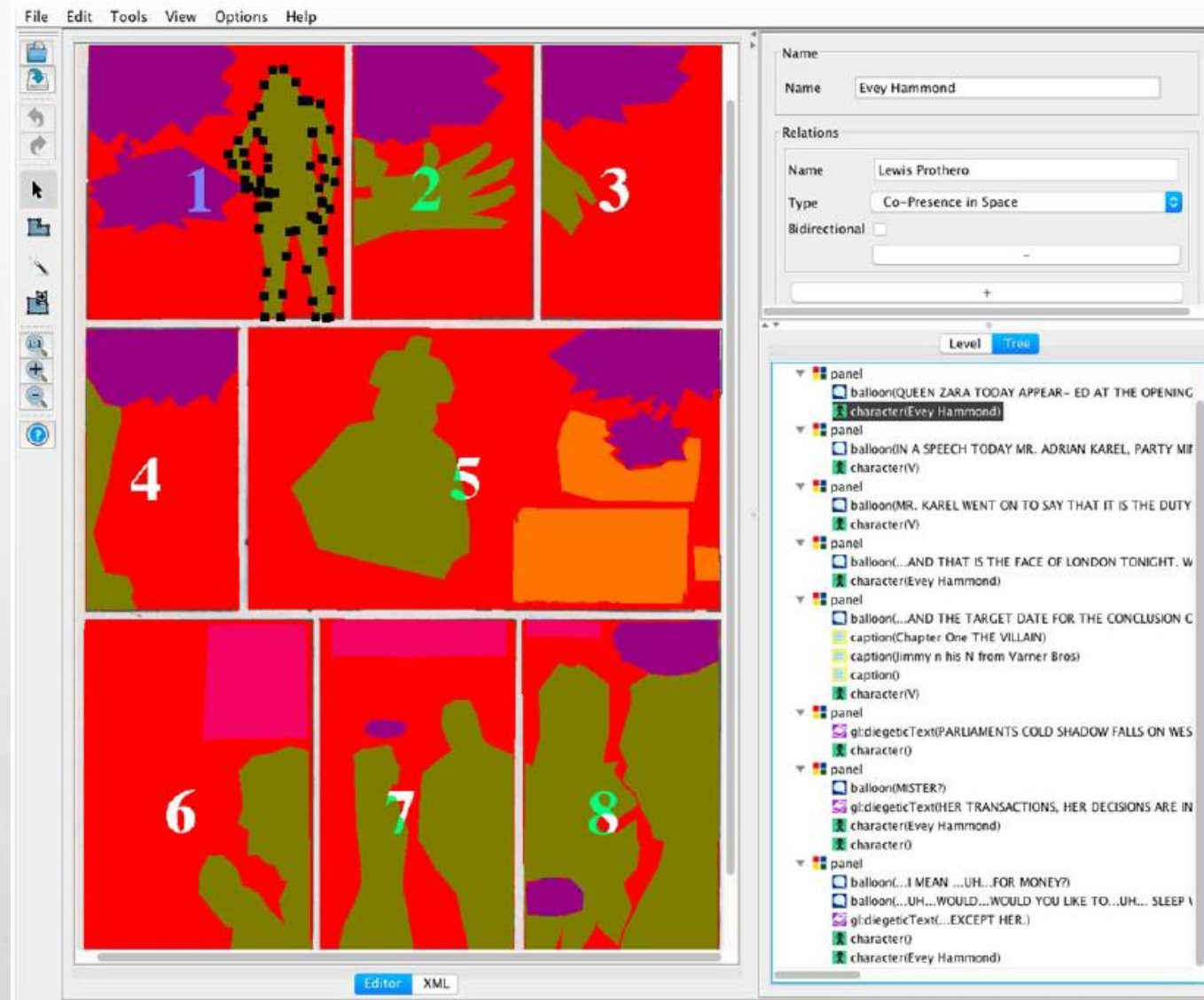
Ethnicity per Source (US Publications)



**DISTRIBUTION OF GENDER AND ETHNICITY OF US TITLES
PER SOURCE**



GNML (GRAPHIC NARRATIVE MARKUP LANGUAGE)



M3-EDITOR INTERFACE

ПЛЮСЫ И МИНУСЫ

ПЛЮСЫ

1. БАЛАНС И РЕПРЕЗЕНТАТИВНОСТЬ
2. ПОДХОДИТ КАК ДЛЯ ГУМАНИТАРНЫХ ИССЛЕДОВАНИЙ (В Т.Ч. КОГНИТИВИСТИКИ), ТАК И В КАЧЕСТВЕ ДАТАСЕТА ДЛЯ КОМПЬЮТЕРНЫХ НАУК
3. ЕСТЬ EYE-TRACKING ДАННЫЕ !
4. ХОРОШО ПРОДУМАНА АННОТАЦИЯ И УДОБНО РАЗМЕЧАТЬ
5. ХОТЯТ РАСШИРЯТЬ КОРПУС ДАЛЬШЕ, В 2018 ГОДУ – 250 TITLES

МИНУСЫ

1. ТОЛЬКО ПЕРВЫЕ 10 СТРАНИЦ ИЛИ 1 ГЛАВА
2. OCR НА ОСНОВЕ TESSERACT РАБОТАЕТ НЕ ОЧЕНЬ И НАДО ЕЩЕ ЕГО ОБУЧИТЬ
3. ПРОБЛЕМЫ С COPYRIGHT PROTECTION => СМОГУТ ДЕЛИТЬСЯ ТОЛЬКО С ОГРАНИЧЕННЫМИ ГРУППАМИ ДРУГИХ УЧЕНЫХ (НЕМЕЦКИЕ ЗАКОНЫ ОНИ ТАКИЕ)
4. НЕТ ИНФОРМАЦИИ О ГР: ОНИ В ОРИГИНАЛЕ НА АНГЛИЙСКОМ ИЛИ ПЕРЕВОДЫ ТОЖЕ БРАЛИ?

Вопрос: как быть с переводами? Очень много графических рассказов переводят на английский

ПСС, ПАРЕНЬ



**РАЗМЕТИТЬ КОРПУС НЕ
ХОЧЕШЬ?**

captions.ru
risovach.ru

**СПАСИБО ЗА
ВНИМАНИЕ!**



СПИСОК ЛИТЕРАТУРЫ:

- **A. FUJIMOTO ET AL., “MANGA109 DATASET AND CREATION OF METADATA,” IN PROCEEDINGS OF THE 1ST INTERNATIONAL WORKSHOP ON COMICS ANALYSIS, PROCESSING AND UNDERSTANDING - MANPU '16, 2016, PP. 1–5.**
- **M. IYYER ET AL., “THE AMAZING MYSTERIES OF THE GUTTER: DRAWING INFERENCES BETWEEN PANELS IN COMIC BOOK NARRATIVES,” IN IEEE CONFERENCE ON COMPUTER VISION AND PATTERN RECOGNITION, 2016.**
- **C. GUERIN ET AL., “EBDTHEQUE: A REPRESENTATIVE DATABASE OF COMICS,” IN PROCEEDINGS OF THE INTERNATIONAL CONFERENCE ON DOCUMENT ANALYSIS AND RECOGNITION, ICDAR, 2013, PP. 1145–1149.**