



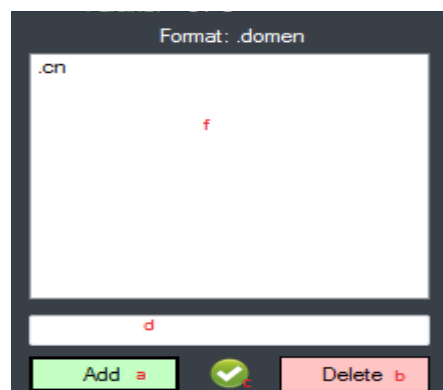
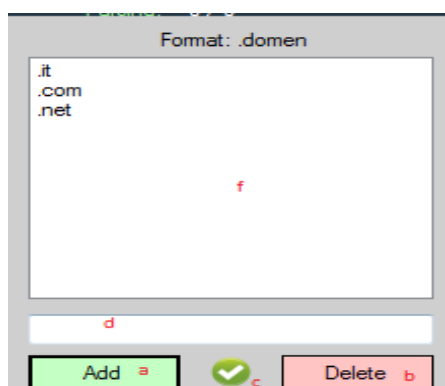
- 1-5 Кнопки управления
1. Загрузить лист с запросами.
2. Пауза.
3. Загрузить лист с прокси.
4. Стоп.
5. Начать работу.
6. Выбрать кол-во потоков. Мин. 1, макс. 600
7. Выбрать таймаут подключения. Мин 1000мс, макс. 25000мс.
8. Выбрать тип прокси: Не использовать, HTTP, Sock4, Socks5
9. Подтвердить функцию прокси по ссылке.
10. Поле для ссылки с прокси.
11. Время обновления прокси(в минутах)
12. Поле вывода информации для пользователей.
13. Кол-во страниц для парсинга.
14. Если вкл, то создается 1 файл.

One file
 Result.txt

Если выкл, то будут созданы файлы с результатом в зависимости от выбранного поисковика.

One file
 Bing.txt
 Yahoo.txt
15. Подать звуковой сигнал по завершению.
16. Сохранять ссылки только с знаком «?»
17. Не сохраняет ссылки формата: "http://site.com/", только "<http://site.com/aaa>", если выкл, сохраняет оба варианта..

18 — 19. Белый и черный список. Писать без точки!!!



18. Белый список. Сохранены будут только домены включенные в этот список.

- a – Добавить домен.
- b – Удалить выбранный домен.
- c – Выйти/Подтвердить
- d – Поле ввода домена(Без точки), к примеру просто it или com
- f – Список включенных доменов

19. Черный список. Домены включенные в этот список не будут сохранены.

- a – Добавить домен.
- b – Удалить выбранный домен.
- c – Выйти/Подтвердить
- d – Поле ввода домена(Без точки), к примеру просто it или com
- f – Список включенных доменов

20. Функция ISearch

При вкл.(возможное снижение скорости до 10% на больших объемах):

Только в режиме «Manu»/ Ссылки одного поисковика, после перехода к другому не будут добавлены и рассматриваются как дубликаты.

При выкл.(Максимальная скорость парсинга):

Одинаковые ссылки из одного и второго поисковика будут сохранены.

К примеру: Парсится Google и Bing, ссылка xxx при нахождение в Google сохраняется, а так же при нахождении в Bing. При вкл режиме, будут сохранятся в памяти для выевления дублей, а это снижает скорость, особенно на больших объемах 10кк+

21-24 Удаление дублей.

21. Режим string. Удаление дубликатов строки

Вход:

```
"https://site.com/game.php?id=10270"  
"https://site.com/game.php?id=10270"  
"https://site.com/game.php?id=62585"  
"https://site.com/thread.php?cat=2452"  
"https://m.site.com/mains/hots"
```

Сохранить:

```
"https://site.com/game.php?id=10270"  
"https://site.com/game.php?id=62585"  
"https://site.com/thread.php?cat=2452"  
"https://m.site.com/mains/hots"
```

22. Режим to= . Удаление дублей до знака "="

Вход:

```
"https://site.com/game.php?id=10270"
```

"https://site.com/game.php?id=10270"
"https://site.com/game.php?id=62585"
"https://site.com/thread.php?cat=2452"
"https://m.site.com/mains/hots"

Сохранить:

"<https://site.com/game.php?id=10270>"
"https://site.com/thread.php?cat=2452"
"https://m.site.com/mains/hots"

если нету знака "=" - работает в режиме string, удалить дубликаты строки.

23. Режим site - Удалить дубликаты сайтов

Вход:

"https://site.com/game.php?id=10270"
"https://site.com/game.php?id=10270"
"https://site.com/game.php?id=62585"
"https://site.com/thread.php?cat=2452"
"https://m.site.com/mains/hots"

Сохранить:

"<https://site.com/game.php?id=10270>"
"<https://m.site.com/mains/hots>"

24. Режим site++ - Удалить дубликаты сайтов по названию.

Вход:

"https://site.com/game.php?id=10270"
"https://site.com/game.php?id=10270"
"https://site.com/game.php?id=62585"
"https://site.com/thread.php?cat=2452"
"https://m.site.com/mains/hots"

Сохранить:

"<https://site.com/game.php?id=10270>"

Тесты удаление дубликатов. 1к запросов:

Поисковик тот же, настройки те же.
Спарсилось/сохр.

Parsing: 0 / 1000
Proxy: 0 string
Link: 49976/28789

Parsing: 0 / 1000
Proxy: 0 to=
Link: 49955/25450

Parsing: 0 / 1000
Proxy: 0 site
Link: 49975/17003

Parsing: 0 / 1000
Proxy: 0 site+
Link: 49976/14719

25. Последняя версия.

26-32 Счетчики

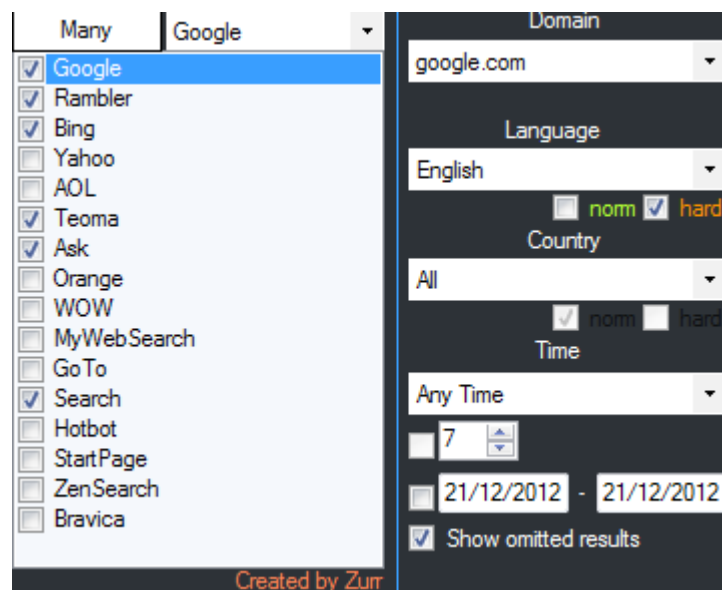
- 26. Счетчик листа с запросами: **Обработано/Всего**
- 27. Счетчик прокси. **Используется**
- 28. Счетчик парсинга: **Всего/Сохранено**
- 29. Счетчик очереди Поисковых Систем: **Выполняется/Всего**
- 30. Счетчик ошибок подключения.
- 31. С какой поисковой системы идет парсинг.
- 32. Счетчик потоков.

33 — Кнопки формы

- 33. Посмотреть информацию о программе.
- 34. Открыть список изменений.
- 35. Открыть папку с результатом
- 36. Свернуть в трей
- 37. Закрыть форму

38. Одиночный поиск.

39. Кнопка «Many» - создать очередь из поисковых систем.

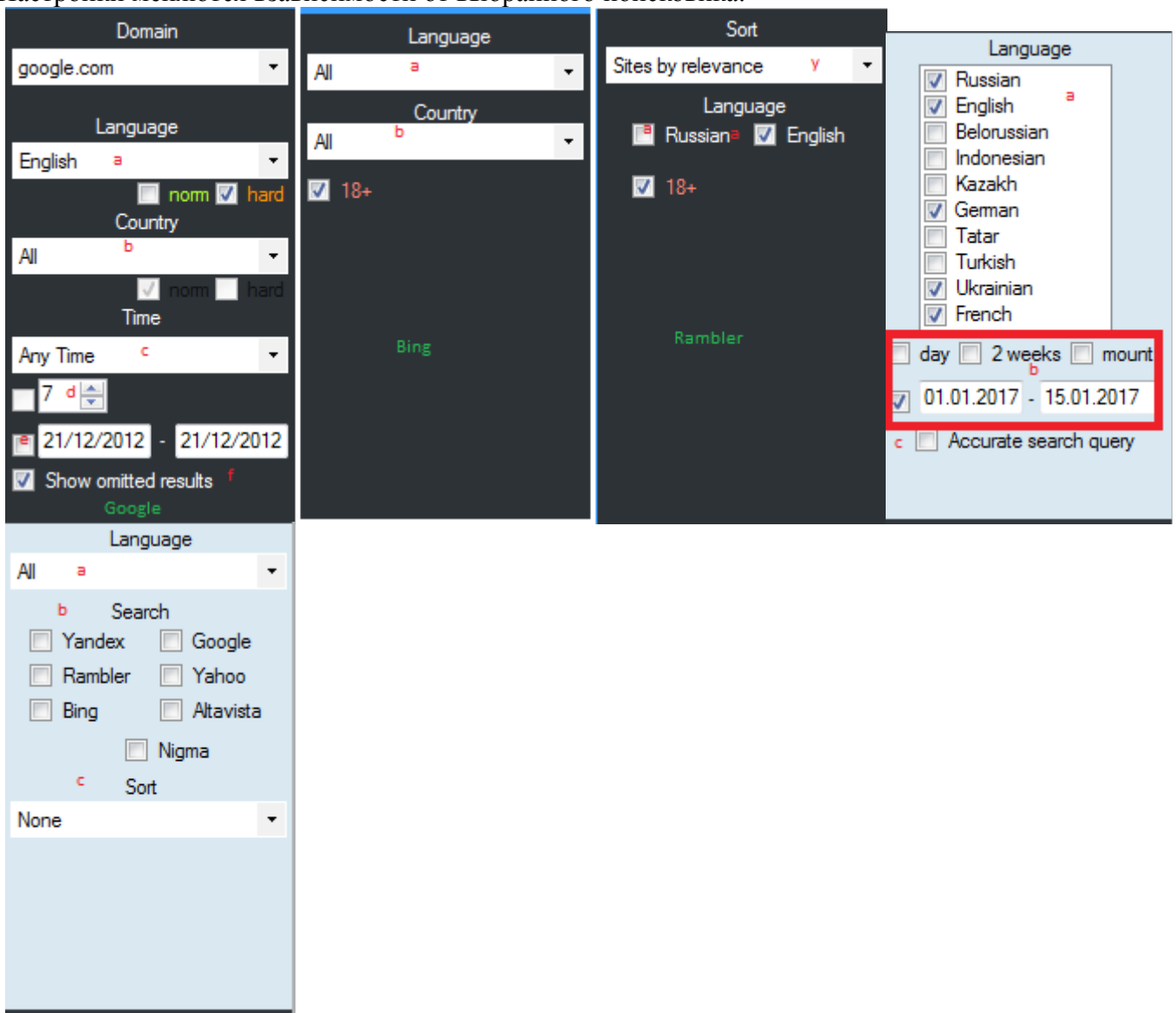


Метод работы тот же, нужные отметить галочкой, если для отмеченного поисковика имеются дополнительные настройки — они откроются.

40. Поле с информацией о поисковой системе.

41. Дополнительные/Индивидуальные настройки поисковой системы.

Настройки меняются в зависимости от выбранного поисковика:



Google:

Domain – выбрать домен поисковой системы.

a – Выбрать язык.

Norm – умеренно.

Hard – максимально.

b – Выбрать страну.

Norm – умеренно.

Hard – максимально.

c – Указать период времени.

d – если не Any Time, то для hour(часов) работает как 7 часов, для минут как 7 минут и т.д

e – Диапазон времени в формате dd/mm/yyyy – dd/mm/yyyy

f – Показывать скрытые результаты

Bing:

a – Выбрать язык.

b – Выбрать страну.

18+ - Показать скрытый результат.

Rambler:

y – сортировка результата

Pages on relevance - Страницы по релевантности.

Sites by relevance - Сайты по релевантности.

Pages by date - Страницы по дате.

Sites by date - Сайты по дате.

a – Выбрать язык.

Russian – русский.

English – английский.

18+ - Показать скрытый результат.

36. Выбор кол-во страниц для парсинга.

37. Выпадающий лист с поисковыми системами.

Yandex:

a – выбор языка.

b – выбор времени парсинга.

c – точный запрос.

Nigma:

a – выбор языка.

b – выбор поисковых систем.

c – Выбор сортировки по дате(новые/старые)