WIKIRACER

FLY THROUGH WIKI

AGENDA

Code

Examples (Visuals)

Final conclusion & takeaways

THE POWER OF

TTD
(CODE)

UNIT TESTS

TAKEAWAYS:
- ANYHOW
- WIREMOCK
- LOG2
- CLAP

👤 Walentalien

```rust
#[cfg(test)]
mod tests {
    use ...;
    // tests for construct_url start here
```

👤 Walentalien

```rust
    #[test]
    fn test_get_url() -> Result<(), Box<dyn std::error::Error>> {
        let root_url :Url = Url::parse( input: "https://google.com")?;
        let another_full_url :Result<Url, ParseError> = Url::parse( input: "https://anot
        let result_url :Url = construct_url(another_full_url.clone().unwrap().as
        assert_eq!(result_url, another_full_url.unwrap());
        Ok(())
    }
}
```

👤 Walentalien

```rust
    #[test]
    fn test_get_url_with_relative_path() -> Result<(), Box<dyn std::error::Error
        let root_url :Url = Url::parse( input: "https://groq.xyz")?;
        // for the path arg we have relative path
        //let relative_path = Url::parse("/some/relative/path")?;
        let relative_path :&str =  "/some/relative/path";
        let result_url :Url = construct_url(relative_path, root_url )?;
        let expected_url :Url = Url::parse( input: "https://groq.xyz/some/relative/
        assert_eq!(result_url, expected_url);
        Ok(())
    }
}
```

crawler.rs   scraped_links/log.txt   scraped_links/scrape_page_log.txt   log.txt   scrape_page_log.txt   .gitignor

```rust
391          mod tests {
                  👤 Walentalien
488          #[tokio::test]
489          async fn test_scrape_empty_page() -> Result<(), Box<dyn std::error::Error>> {
490              let client: Client = reqwest::Client::new();
491              //let url = Url::parse("https://example.com/empty")?;
492              let config: CrawlerConfig = CrawlerConfig {
493                  starting_url: Url::parse( input: "https://localhost")?,
494                  scraping_foreign_hosts: false,
495                  max_urls: 0,
496                  max_depth: 0,
497                  thread_count: 0,
498                  request_delay_ms: 0,
499                  max_retries: 0,
500                  request_timeout_sec: 2,
501                  target_url: None,
502              };
503
504              let mock_server: MockServer = wiremock::MockServer::start().await;
505              Mock::given(method("GET"))
506                  .and(path("/empty")): MockBuilder
507                  .respond_with(ResponseTemplate::new( s: 200).set_body_string("")): Mock
508                  .mount(&mock_server): impl Future<Output=()>+Sized
509                  .await;
510              let url: Url = Url::parse( input: &format!("{}/empty", &mock_server.uri()))?;
511              let result: Result<HashSet<Url>> = scrape_page(url, &client, &config).await;
512              assert_eq!(result?, HashSet::<Url>::new());
513              Ok(())
514          }
515
tests
```

```rust
2 usages    ≗ Walentalien *
#[macro_export]
macro_rules! write_url {
    // Case: HashSet or iterable
    ($path:expr, [$($url:expr),+ $(,)?]) => {{
        #[cfg(debug_assertions)] // doesnt run in release
        {
            use std::fs::OpenOptions;
            use std::io::Write;
            use std::time::Duration;

            let elapsed = $crate::START_TIME.elapsed();
            let minutes = elapsed.as_secs() / 60;
            let seconds = elapsed.as_secs() % 60;

            let result = OpenOptions::new()
                .create(true)
                .append(true)
                .open($path)
                .and_then(|mut file| {
                    $(
                        writeln!(file, "[{:02}:{:02}] {}", minutes, seconds, $url)?;
                    )+
                    Ok(())
```

OPEN FOR CONTRIBUTIONS

Contribution Guideline

Open for Contribution Looking for help with:

- GUI Application

EXAMPLES
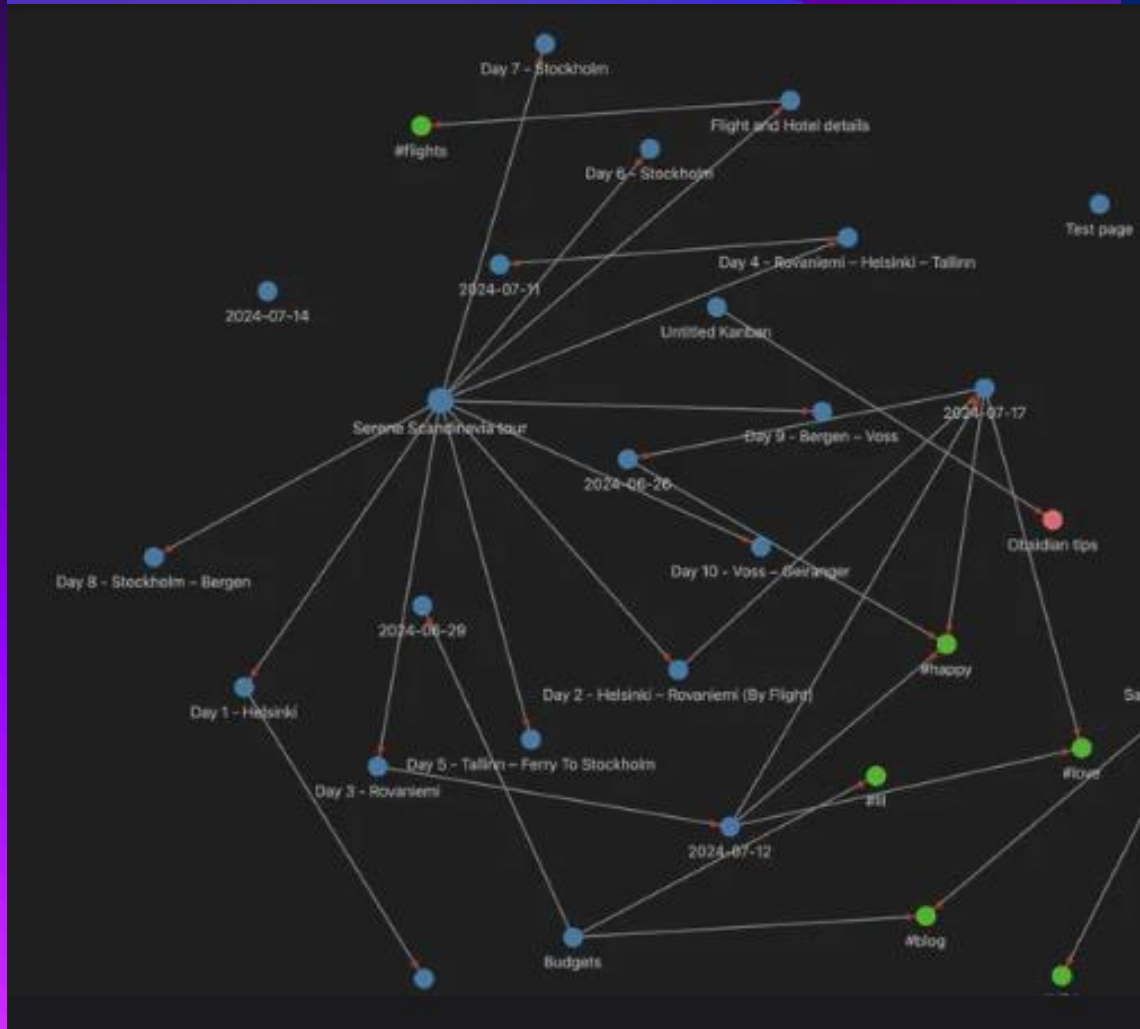
```
Terminal    Local  ×   +  ∨                                                                              ⋮  —

[2025-06-17 15:35:45.274] [DEBUG] [WikiRacer::crawler:255] Skipped foreign host link: https://en.m.wikipedia.org/w/index.php?title=Pope_John_XIV&mobileaction=toggle_view_mob
ile
[2025-06-17 15:35:45.274] [DEBUG] [WikiRacer::crawler:255] Skipped foreign host link: https://www.wikimedia.org/
[2025-06-17 15:35:45.274] [DEBUG] [WikiRacer::crawler:255] Skipped foreign host link: https://www.mediawiki.org/
[2025-06-17 15:35:45.275] [INFO] [WikiRacer::crawler:266] Found 848 urls on page https://en.wikipedia.org/wiki/Pope_John_XIV
[2025-06-17 15:35:45.353] [INFO] [WikiRacer::crawler:313] Worker 3: Max URLs reached
[2025-06-17 15:35:45.353] [INFO] [WikiRacer::crawler:369] Worker 3 finished
[2025-06-17 15:35:45.378] [INFO] [WikiRacer::crawler:313] Worker 2: Max URLs reached
[2025-06-17 15:35:45.378] [INFO] [WikiRacer::crawler:369] Worker 2 finished
[2025-06-17 15:35:45.378] [INFO] [WikiRacer:58] Crawling completed. Total links found: 1002
[2025-06-17 15:35:45.634] [DEBUG] [WikiRacer::pathfinder:13] Searching for path from https://en.wikipedia.org/wiki/Pope to https://en.wikipedia.org/wiki/Banana
[2025-06-17 15:35:45.635] [DEBUG] [WikiRacer::pathfinder:14] Graph contains 1002 nodes
[2025-06-17 15:35:45.644] [DEBUG] [WikiRacer::pathfinder:43] Searching at depth 1, 2001 nodes in queue
[2025-06-17 15:35:47.214] [DEBUG] [WikiRacer::pathfinder:43] Searching at depth 2, 160250 nodes in queue
[2025-06-17 15:35:47.414] [DEBUG] [WikiRacer::pathfinder:43] Searching at depth 3, 0 nodes in queue
[2025-06-17 15:35:47.415] [DEBUG] [WikiRacer::pathfinder:73] No path found after searching 162253 nodes
[2025-06-17 15:35:47.651] [INFO] [WikiRacer:69] No path found between https://en.wikipedia.org/wiki/Pope and https://en.wikipedia.org/wiki/Banana
walentyn@lenovo:/media/walentyn/HDD/MY_GITHUB_REPOS/WikiRacer$ █
```

DEVELOPMENT

Your ability to communicate effectively will leave a lasting impact on your audience

Effectively communicating involves not only delivering a message but also resonating with the experiences, values, and emotions of those listening

SPEAKING ENGAGEMENT METRICS

From	To	Links between
Prezydent Mozabbiku	AGH	2
Biedronka	Korporacja	1
C	Rust	0
CCP	Poland	1
CCP	Dumplings	2

t path (2 steps):
https://en.wikipedia.org/wiki/Chinese_Communist_Party
https://en.wikipedia.org/wiki/Radio_Free_Europe/Radio_Liberty
https://en.wikipedia.org/wiki/Poland
6-17 16:01:21.995] [WikiRacer:64] Path found!
1.075] [WikiRacer:64] Path found!
1.075] [INFO] [WikiRacer:66] Number of links between pages:
0.851] [DEBUG] [WikiRacer::pathfinder:54] Found target at depth 3
6-17 16:01:22.090] [INFO] [WikiRacer:66] Number of links between pages: 3
eps):
•wikipedia.org/wiki/Chinese_Communist_Party
•wikipedia.org/wiki/Republic_of_China_Armed_Forces
•wikipedia.org/wiki/Taiwanese_cuisine
•wikipedia.org/wiki/Dumpling
1.075] [INFO] [WikiRacer:66] Number of links between pages: 3

THANK YOU

Link to repo ->>>>