Práce se zabývá tvorbou webového robota. Jeho úkolem je rekurzivně stahovat z internetu české stránky a čistit je na samotný prostý text (žádné HTML značky, styly nebo skripty). Ten potom bude využit pro tvorbu obrovského jazykového korpusu, užitečného pro další výzkum. Klíčovou vlastností robota je nenápadnost běhu, nezatěžování cizích prostředků a plné respektování nezávazného doporučení Robots Exclusion Standard. Robot je napsán v jazyce Python a intenzivně využívá jeho standardní knihovny a rychlou práci s textovými řetězci. Vzhledem k charakteru úlohy jsme se rozhodli pro paralelní implementaci, která by měla plně využít šířku pásma. S tímto záměrem jsme měli úspěch. Výsledkem práce je tedy robot připravený získat dostatek textů pro korpus. Samozřejmě je ale použitelný i pro jiné účely, zvlášť tam, kde je potřeba šetrnost k cizím prostředkům. Kromě jeho přínosu pro lingvistiku poskytuje i zajímavé informace o obsahu českého internetu.