

УДК 004.73; 004.77

Ахмедьянов В. – ст. гр. СНм-52

Тернопільський національний технічний університет імені Івана Пулюя

ВИДИ РЕСУРСІВ «ГЛИБИННОГО ВЕБ»

Науковий керівник: к.т.н., доцент Мацюк О.В.

Ahmedyanov V.

Ternopil Ivan Pul'uy National Technical University

TYPES OF RESOURCES "DEEP WEB"

Supervisor: Matsiuk O.

Ключові слова: ресурс, «глибинний веб»

Keywords: resource, "Deep web"

Існує кілька типів ресурсів глибинного веб, наприклад, це можуть бути веб-сторінки, які швидко старіють. Крім того, до глибинного веб відносяться веб-ресурси, що представляють собою мультимедійну інформацію. Як відомо, в даний час ще не існує задовільних алгоритмів пошуку НЕ текстової інформації. Динамічно генеровані за запитом сторінки також часто потрапляють в глибинний веб. Найчастіше без запиту таких сторінок не існує, вони генеруються при запиті до баз даних. Виходить, що інформація начебто і є присутня у веб-просторі, але виникає вона лише в момент обробки запиту, а універсального алгоритму заповнення їх роботами пошукових форм не існує. І, нарешті, якщо на веб-ресурс не ведуть ніякі посилання, то роботи пошукових систем ніяким чином не можуть дізнатися про його існування.

Засновник компанії BrightPlanet Майкл Бергман (Michael K. Bergman) виділив 12 різновидів глибинних веб-ресурсів, що відносяться до класу онлайн-баз даних. У списку опинилися як традиційні бази даних (патенти, медицина і фінанси), так і публічні ресурси – оголошення про пошук роботи, чати, бібліотеки, довідники. Бергман зарахував до глибинних ресурсів і спеціалізовані пошукові системи, які обслуговують певні галузі або ринки, бази даних яких не включаються в глобальні каталоги традиційних пошукових служб.

До глибинного веб також відносяться численні системи інтерактивної взаємодії з користувачами – допомоги, консультування, навчання, що вимагають участі людей для формування динамічних відповідей від серверів. До них також можна віднести і закриті (повністю або частково) інформацію, доступну користувачам мережі Інтернет тільки з певних адрес, груп адрес, іноді міст або країн. До «прихованої» частини мережі більшість зараховує і веб-сторінки, зареєстровані на безкоштовних серверах, які індексуються, в кращому випадку, лише частково – пошукові системи, щоб уникнути рекламного спаму не прагнуть обходити у повному обсязі.

До глибинного веб також відноситься категорія так званих «сірих» сайтів, що функціонують на основі динамічних систем керування контентом (Dynamic Content Management Systems). В пошукових системах зазвичай обмежується глибина індексування таких сайтів, щоб уникнути можливого циклічного перегляду одних і тих же сторінок.