



В основу работы положена идея о том, что все составляющие элементы схемы, корневые морфемы и их устойчивые сочетания, соответствующие базовым понятиям предметной области, а также флексии и устойчивые конструкции семантического синтаксиса, характеризующие типы отношений между объектами, выражаемыми понятиями, могут быть выделены на основе статистического анализа.

Связи между понятиями семантической сети могут быть установлены по их совместному появлению в рамках структурных единиц текста, в простейшем случае - в предложениях. Сформированная сеть позволяет реализовать анализ с точки зрения выделения в тексте ключевых понятий, СФЕ и установления связей между ними.

Проще говоря, проанализировав содержимое отдельного документа или целого информационного пакета, программа самостоятельно найдет наиболее значимые понятия и сформирует из них семантическую сеть. Подобная структура полностью характеризует содержание обрабатываемых данных, причем каждый элемент семантической сети (или понятие) описывается своим весом и набором связей с другими элементами - контекстным узлом.

Вес элемента характеризует относительную смысловую значимость выраженной им темы во всем множестве текстов по сравнению со значимостью других и принимает числовые значения в диапазоне от единицы до ста.