Тезаурус – словарь, охватывающие понятия, определения и термины специальной области знаний. Слова в тезаурусах упорядочены по смысловой близости, не по алфавиту.

Наиболее распространёнными типами смысловых отношений между словами в тезаурусах являются:

* синонимия, базирующаяся на критерии, что два выражения являются синонимичными, если замена одного из них на другое в предложении не меняет значения истинности этого высказывания, например, быстрый – шустрый, бортпроводница – стюардесса;
* антонимия, основанная на смысловом противопоставлении, например, тёплый – холодный, светло – темно;
* гипо-гиперонимия, представляющая собой отношение общего и частного, например, машина – самосвал;
* меронимия, т.е. отношение часть-целое, например, компьютер – процессор, тетрадь – страница.

Синсетом называется множество слов, связанных отношением синонимии. Синсеты разбивают множество всех лексических единиц на классы эквивалентности. Если для некоторого слова не существует синонимов, то соответствующий ему синсет будет состоять только из одного слова. При работе со словом учитываются все его значения, особенно те, в которых это слово является синонимом к другим словам. Многозначные слова, рассматриваемые в разных значениях, входят и в разные синсеты: золотая (монета) – сделанная из золота и золотой (работник) – хороший.

WordNet - это огромная лексическая база знаний для английского языка. WordNet является семантической сетью, узлы которойпредставляют собой синсеты, связанные различными отношениями, такими как гипонимия, гиперонимия, голонимия, меронимия и т.п

WordNet приобрёл популярность благодаря его содержательными и структурными характеристиками. Принстонский WordNet и все последующие варианты для других языков направлены на отображение состава и структуры лексической системы языка в целом, а не отдельных тематических областей. Для каждого синсета имеется описание на естественном языке, а так же примеры использования входящих в него слов. В состав тезауруса входят лексемы, относящиеся к четырем частям речи: прилагательное, существительное, наречие и глагол. Лексемы различных частей речи хранятся отдельно, и описания, соответствующие каждой части речи, имеют различную структуру.

Существительные, прилагательные, глаголы, наречия сгруппированы в наборы когнитивных синонимов (синсеты), каждый из которых выражает отдельное значение. Синсеты взаимосвязаны между собой посредством концептуально-семантических и лексических отношений. WordNet является свободно распространяющейся и соответственно общедоступной для загрузки базой знаний. Тем самым структура WordNet делает его полезным инструментом для вычислительной лингвистики и обработки естественного языка.

Существительные, прилагательные, глаголы, наречия сгруппированы в наборы когнитивных синонимов (синсеты), каждый из которых выражает отдельное значение. Синсеты взаимосвязаны между собой посредством концептуально-семантических и лексических отношений. WordNet является свободно распространяющейся и соответственно общедоступной для загрузки базой знаний. Тем самым структура WordNet делает его полезным инструментом для вычислительной лингвистики и обработки естественного языка.

Слова в WordNet группируются вместе на основе их значений. WordNet внешне напоминает тезаурус, однако есть некоторые важные различия. Во-первых, WordNet связывает не только словоформы, но и слова со схожим смыслом. Во-вторых, WordNet отмечает семантические отношения между словами, тогда как группировки слов в тезаурусе не следуют какой-либо явной схеме, кроме сходства.

Основным отношением между словами в WordNet является синонимия. Синонимы - слова, которые обозначают одну и ту же концепцию и являются взаимозаменяемыми во многих контекстах. Они группируются в неупорядоченные наборы (синсеты). Каждый из 117 000 синсетатов WordNet связан с другими синсетами с помощью небольшого числа смысловых отношений. Кроме того, синсет содержит краткое определение и, в большинстве случаев, одно или несколько коротких предложений, иллюстрирующих использование слов из данного синсета. Формы слов с несколькими различными значениями представлены в виде множества различных синсетов.

Синонимы обязаны быть взаимозаменяемы хотя бы в некотором непустом множестве контекстов. Для отношения синонимии не требуется заменимость всех синонимов во всех контекстах, иначе в естественном языке было бы слишком мало синонимов. Для существительных в WordNet установлены следующие семантические отношения: синонимия, антонимия, гипонимия/гиперонимия, меронимия.

Наиболее часто встречающимся отношением между синсетами является гиперонимия и гипонимия. Гиперонимия связывает более общие синсеты, такие как мебель, с более специфическими, такими как кровать. Таким образом, согласно WordNet в категорию мебели входит кровать, которая, в свою очередь, включает в себя двухъярусную кровать. Наоборот, понятия типа кровати и двухъярусной кровати составляют категорию мебели. Все иерархии существительных в конечном счете поднимаются на корневой узел. Отношение гипонимии является переходным: если кресло является своего рода стулом, а стул есть мебель, то кресло является своего рода мебелью. Синсет А – гипоним синсета В, в том случае, когда существуют предложения типа А есть (является разновидностью) В. И соответственно гаоборот Синсет А – гипероним синсета В, в том случае, когда существуют предложения типа А имеет разновитность В.

Меронимия или другими словами отношение «часть-целое» имеет место между синсетами, такими как, например, стул и спинка, стул и ножки. В WordNet выделяются три подвида отношения часть-целое: быть частью, быть элементом, быть сделанным из. Части у различных сущностей могут иметь одинаковое название, например, острие может быть у иголки, карандаша, стрелы, ножа, булавки и т.д. Таким образом А является меронимом В в том случае, если предложения вида А содержит В и А является частью В являются естественными для А и В, интерпретируемых как родовые понятия.

Прилагательные подразделяются на относительные и качественные. В связи с тем, что человек при описании качественного прилагательного зачастую называет антонимы для него, семантическое описание качественных прилагательных основано на отношении антонимии. Пары «прямых» антонимов, таких как, например, сухой и мокрый, пожилой и молодой отражают смысловое значение друг друга. Таким образом семантически связанные прилагательные являются «косвенными антонимами». В отдельную группу прилагательных выносятся прилагательные оттенков цветов. Значение относительных прилагательных является отсылкой к соответствующему синсету существительных, таким образом относительные прилагательные указывают на существительные от которых они образуются.

Так же в WordNet выделяют 2 категории глаголов согласно их смысловому значению: глаголы, обозначающие действия (действия и события), и глаголы состояния. Среди глаголов действий и событий выделяют следующие 14 групп: контакта, движения, коммуникации, восприятия, изменения, соревнования, познания, создания, эмоций, потребления, обладания, ухода за телом и глаголы, относящиеся к социальному поведению. Однако, в связи с тем, что нельзя однозначно отнести многие глаголы к той или иной группе, границы между группами точно не установлены. Отношение логического следования устанавливается между синсетами глаголов А и В, если из предложения “…выполняется А" следует, что “… выполняется В”. Например, из того, что "Девушка говорит", следует, что "Девушка издаёт звуки".

Для установления иерархических отношений между глаголами было введено отношение тропонимии. То есть "… делать А есть делать В в определённой форме". Например, “Шептать – это тихо разговаривать". Отношение тропонимии – особый вид отношения следования. Отношение причины связывает два глагольных синсета, один из которых называется результатив, а второй каузатив. Отношение причины также может быть рассмотрено как специальный случай следования. Если А влечёт за собой В, то из В также логически следует А.

Большинство отношений WordNet связывают слова, являющиеся одной частью речи. Таким образом можно сказать, что WordNet действительно состоит из четырех подсетей, по одному для существительных, глаголов, прилагательных и наречий, с несколькими перекрестными POS-указателями.