

Thesaurus Design (Group Assignment No. 1)
LIBR 247
Dr. Weedman

Kristin Yiotis

Introduction

The purpose of the following thesaurus is to provide terms of use to those who will search our group's newly constructed database of articles. This database currently contains 22 articles all within the broad field of outdoor, non-organized sports. The descriptors were specifically chosen for the benefit of a non-specialized user-base, since these articles are of interest to general users. However, we acknowledge that outdoor sports enthusiasts and specialists will more than likely be its primary users. [If you are creating a database for enthusiasts and specialists, why did you choose terms for the benefit of non-specialists? \(It's not clear to me whether your user group of non-specialists includes enthusiasts or not.\) I'm not sure, actually, how much difference there would be between vocabularies created with the two different groups in mind, but assuming there is a difference, then designing terms for a group that you don't anticipate will be your primary users seems like it might not be a very good idea.](#)

This thesaurus has been created with a "flat" architecture in mind: there are 33 bushes of terms, with some of these bushes consisting of a single term. The hierarchies in these bushes are therefore simple for the user to follow, and the structural design allows for easy expansion of the thesaurus as the database size increases. This also complies with the subject matter. Although a majority of the 287 total terms in this thesaurus are used to identify outdoor sports articles, there are fringe areas such as Bodybuilding, Care

did you mean car? maintenance, and the technical aspects of some of the sports that defy easy categorization or fixture into a sensible hierarchy with the other terms.

The standard relationships of broader term (BT), narrower term (NT), related term (RT), used for (UF), and USE have been identified within the alphabetical hierarchy.

We have attempted at each step in building this thesaurus to comply with NISO (www.niso.org) guidelines concerning monolingual thesaurus construction. One of the NISO standards concerning introductions includes a recommendation to supply separate numbers for entry terms and descriptors. This thesaurus contains eight equivalent terms (UF or USE) out of 287 descriptors. While this is a low number for entry terms, descriptors were selected for their simplicity. An increase in technical language, which may or may not occur with additions to the database, would undoubtedly lower this ratio. In other words, an increase in technical language would lower the number of entry terms in relationship to the number of descriptors? (That's what this sentence literally says, but its structure made me first think it meant the opposite.) It could work either way – technical terms could increase the precision of the language, thus requiring fewer near synonyms, or it could create the need for more near synonyms to lead the user from a common term to a more technical term.

This thesaurus will be updated as necessitated by the addition of articles to the database. Queries concerning these updates can be directed to the students listed at the top of this page (info@get-outdoors.com).

Thesaurus

Bush display

Barrel racing techniques

Boatbuilding techniques

Flag making

Boating

Sailing

Night sailing

Yachting

Boating equipment

Anchors

Boats

Boat trailers

Flags

Sails

Boating techniques

Navigating This worries me a little because it isn't completely inside the boundaries of its BT – navigating is done for other activities besides boating – for instance, hang gliding and hiking.

Bodybuilding

Nutrition training

Strength training

Car maintenance techniques

Cycling

Bicycle racing

Mountain biking

Motocross racing

Road biking

Cycling equipment

Aluminum frames

Bike racks

Mountain bikes

Road bikes

Cycling techniques

Downhill braking

Golfing

Amateur golfing

Professional golfing

Golf equipment

Golf bags

Golf clubs

Golf shoes

Golf techniques

Swinging This relationship is correct, but the descriptor needs some kind of qualifier to make it clear what it means when the rest of the hierarchy isn't visible.

Horse sports

Barrel racing

Horse-racing

Endurance racing

This relationship isn't correct. Motocross racing involves motorcycle-like vehicles and ATVs, which have engines. A mountain bike is an actual bicycle, one with extra-wide tires and different handle bars. If this is just an indentation error, I'm still not sure it's correct to refer to riding bikes with engines as "cycling."

Harness racing
Thoroughbred racing

Polo

Horse equipment

Bridles

Saddles

Stirrups

Horse training techniques

Horsemanship techniques

Posting

Scuba diving

Technical diving

Wreck diving

Scuba diving equipment

Buoys

Fins

Masks

Tanks

Scuba diving techniques

Buoyancy control

Safety techniques As worded here, this relationship is incorrect. Not all safety

techniques are scuba diving techniques. One way to change it would be to reword it to be “scuba diving safety techniques.” However... It’s more in the spirit of a post coordinate vocabulary to make safety techniques a separate concept, and then allow an indexer to assign it with whatever sport is discussed in an article – you could have Safety techniques and Scuba diving, or Safety techniques and Horsemanship, etc. As the number of magazines and articles grows, you’d be quite likely to need several of those combinations.

Ski equipment

Boots

Skis

Ski racks

Skiing

Cross country skiing

Downhill skiing

Skiing techniques

Balancing These may (or may not) be more specific than needed. However,

Stance as they stand, they need disambiguation. Balance is extremely

important in horseback riding, and “stance” isn’t completely clear in meaning when the word is not in its bush. Some rewording would make them much clearer.

Snowboarding

Snowboarding equipment

Snowboarding techniques

Snowshoeing

Snowshoes

Sports facilities

Several of these terms have meanings outside the scope of diving equipment; qualifiers would disambiguate them.

- Golf courses
- Racetracks
- Regattas [A regatta is an event, a race, not a physical facility.](#)
- Riding stables
- Rodeos [A rodeo is also an event, not a place.](#)
- Swimming pools
- Tennis courts
- Photography equipment
 - Cameras
 - Film
 - Tripods
- Photography of sports
 - Water sports photography
 - Winter sports photography
- Photography techniques
 - Making portfolios
 - Photographing sea creatures
- Vacation places
 - Dude ranches
 - Golf resorts
 - Riding camps
 - Ski resorts
 - Tropical sea resorts

Alphabetical display – BT, NT, RT, UF, and USE.

- Aluminum frames
 - BT Cycling equipment
- Anchors
 - BT Boating equipment
- Balancing
 - BT Skiing techniques
 - BT Snowboarding techniques
- Barrel racing
 - BT Horse sports
 - RT Rodeos
- Barrel racing techniques
- Bicycle racing
 - BT Cycling
- Bike racks
 - BT Cycling equipment
- Bicycling
 - USE Cycling
- Boat trailers
 - BT Boating equipment

- Boatbuilding techniques
 - NT Flag making
- Boating
 - NT Sailing
 - NT Yachting
- Boating equipment
 - NT Anchors
 - NT Boats
 - NT Boat trailers
 - NT Flags
 - NT Sails
- Boating techniques
 - NT Navigating
- Boats
 - BT Boating equipment
- Bodybuilding
 - NT Nutrition training
 - NT Strength training
- Boots
 - BT Ski equipment
- Bridles
 - BT Horse equipment
- Buoyancy control
 - BT Scuba diving techniques
- Buoys
 - BT Scuba diving equipment
- Cameras
 - BT Photography equipment
- Car maintenance techniques
- Cross country skiing
 - BT Skiing
- Cycling
 - NT Bicycle racing
 - NT Motocross racing
 - NT Mountain biking
 - NT Road biking
 - UF Bicycling
- Cycling equipment
 - NT Aluminum frames
 - NT Bike racks
 - NT Mountain bikes
 - NT Road bikes
- Cycling techniques
 - NT Downhill braking
- Deep diving
 - USE Technical diving

- Downhill braking
 - BT Cycling techniques
- Downhill skiing
 - BT Skiing
- Dude ranches
 - BT Vacation places
 - UF Guest ranches
- Endurance racing
 - BT Horse racing
- Equestrian sports
 - USE Horse sports
- Film
 - BT Photography equipment
- Fins
 - BT Scuba diving equipment
- Flag making
 - BT Boatbuilding techniques
- Flags
 - BT Boating equipment
- Golf bags
 - BT Golf equipment
- Golf clubs
 - BT Golf equipment
- Golf courses
 - BT Sports facilities
- Golf equipment
 - NT Golf bags
 - NT Golf clubs
 - NT Golf shoes
- Golf resorts
 - BT Vacation places
- Golf shoes
 - BT Golf equipment
- Golf techniques
 - NT Swinging
- Golfing
 - NT Amateur golfing
 - NT Professional golfing
- Guest ranches
 - USE Dude ranches
- Harness racing
 - BT Horse racing
- Horse care
 - RT Horse training techniques
- Horse equipment
 - NT Bridles

- NT Saddles
- NT Stirrups
- Horse sports
 - NT Horse racing
 - NT Polo
 - NT Rodeos
 - RT Riding sports
 - UF Equestrian sports
- Horse racing
 - BT Horse sports
 - NT Endurance racing
 - NT Harness racing
 - NT Thoroughbred racing
- Horse training techniques
 - RT Horse care
- Horsemanship techniques
 - NT Posting
- Making portfolios
 - BT Photography techniques
- Masks
 - BT Scuba diving equipment
- Motocross racing
 - BT Mountain biking
- Mountain bikes
 - BT Cycling equipment
- Mountain biking
 - BT Cycling
 - NT Motocross racing
- Navigating
 - BT Boating techniques
- Night sailing
 - BT Sailing
- Nutrition training
 - BT Bodybuilding
- Photographing sea creatures
 - BT Photography techniques
- Photography equipment
 - NT Cameras
 - NT Film
 - NT Tripods
- Photography of sports
 - NT Water sports photography
 - NT Winter sports photography
- Photography techniques
 - BT Sports techniques
 - NT Making portfolios

- NT Photographing sea creatures
- RT Photographs
- Photographs
 - RT Photography techniques
- Polo
 - BT Horse sports
- Posting
 - BT Horsemanship techniques
- Professional golfing
 - BT Golfing
- Racetracks
 - BT Sports facilities
- Regattas
 - BT Sports facilities
- Riding camps
 - BT Vacation places
- Riding sports
 - RT Horse sports
- Rigging
 - RT Sails
- Road bikes
 - BT Cycling
- Rodeos
 - BT Sports facilities
 - RT Barrel racing
- Saddles
 - BT Horse equipment
- Safety techniques
 - BT Scuba diving techniques
- Sailing
 - BT Boating
 - NT Night sailing
- Sails
 - BT Boating equipment
 - RT Rigging
- Scuba diving
 - NT Technical diving
 - NT Wreck diving
- Scuba diving equipment
 - NT Anchors
 - NT Buoys
 - NT Fins
 - NT Masks
 - NT Tanks
- Scuba diving techniques
 - BT Sports techniques

- NT Buoyancy control
- NT Safety techniques
- Ski equipment
 - NT Boots
 - NT Skis
 - NT Ski racks
- Ski racks
 - BT Ski equipment
- Ski resorts
 - BT Vacation places
- Skiing
 - NT Cross country skiing
 - NT Downhill skiing
 - RT Winter sports
- Skiing techniques
 - NT Balancing
 - NT Stance
- Skis
 - BT Ski equipment
- Snowboarding
 - RT Winter sports
- Snowboarding equipment
- Snowboarding techniques
 - NT Balancing
- Snowshoeing
 - RT Winter sports
- Snowshoes
- Sports facilities
 - NT Golf courses
 - NT Race tracks
 - NT Regattas
 - NT Riding stables
 - NT Rodeos
 - NT Swimming pools
 - NT Tennis courts
- Stance
 - BT Skiing techniques
- Stirrups
 - BT Horse equipment
- Strength training
 - BT Bodybuilding
 - RT Weightlifting
- Swimming pools
 - BT Sports facilities
- Swinging
 - BT Golf techniques

- Riding stables
 - BT Sports facilities
- Tanks
 - BT Scuba diving equipment
- Technical diving
 - BT Scuba diving
 - RT Wreck diving
 - UF Deep diving
- Tennis courts
 - BT Sports facilities
- Thoroughbred racing
 - BT Horse racing
- Tripods
 - BT Photography equipment
- Tropical sea resorts
 - BT Vacation places
- Vacation places
 - NT Dude ranches
 - NT Golf resorts
 - NT Riding camps
 - NT Ski resorts
 - NT Tropical sea resorts
 - RT Vacations
- Vacations
 - RT Vacation places
- Water sports photography
 - BT Photography of sports
- Winter sports
 - RT Skiing
 - RT Snowboarding
 - RT Snowshoeing
- Winter sports photography
 - BT Photography of sports
- Wreck diving
 - BT Scuba diving
 - RT Technical diving
- Yachting
 - BT Boating

I didn't repeat comments on the correctness of relationships in the alphabetical display. All of the relationships appear to be reciprocal, and to duplicate those shown in the hierarchical display. RTs and entry vocabulary are appropriate.

Discussion

The first step taken once our group pooled our candidate terms was to reconcile the pre-coordinate descriptors and post-coordinate descriptors that our members came up with. “Bodybuilding – Neck,” for example, became “Body building NT Neck exercises” and combine terms in one list. In order to visualize the hierarchy, we developed a tree diagram in which a sport was the top term of a bush with related terms branching down from the top term. Top terms included boating, exercising/body building, cycling, horseback riding, golfing, scuba diving, skiing and photography, as shown in the included .jpeg image. By this process I noticed that nearly each bush had three, second level categories, the name of the sport equipment related to the sport, and techniques or training related to the sport. I simply created an indented list of these bushes and worked out the associative is “associative” really what you mean? Or hierarchical? relationships, some going down five levels. After consulting an expert I realized that the relationship between levels were not all genus-species and part-whole, which were the required relationships, but were mostly topical or related terms. This required us to rethink the entire hierarchy.

The terms with unclear relationships were difficult, and our group had to decide whether or not they were examples of part-whole or instance relationships in order to represent them correctly in our list. Different types of sports equipment, for example, raised questions in our group if they displayed partitive relationships. Can a horse be considered an example of sports equipment, since you can’t ride without one? **Funny!** (But this isn’t a partitive relationship – if you did decide to include horses as sports equipment, it’s genus/species. Partitive relationships involve literal, often physical parts – steering wheels and automobiles, living rooms and houses, Senate Judiciary Committee

and the Senate. Sebastian had difficulty with instance relationships in two articles in particular: “Anchoring” and “1995 Equipment Preview.” The ten specific types of boating anchors (yachtsman, navy, stockless, folding, northill, danforth, plow, grapnel, japan and coral hook) seemed to merit inclusion in the list as instances of the broader term “anchors.” [Actually, these are in a genus/species relationship to “anchors” – they’re still kinds of anchors, not individual anchors.](#) The various examples of snow ski boots (Tecnica 200 AVS, Elan MBX C Series, etc) listed in “1995 Equipment Preview,” however, could also be viewed as instances of that broader term, but the decision was made not to include these terms in our list. This decision was based in part on the conviction that the anchor examples referred to design and construction, while the ski boots examples referred to specific brand names. In other words, the Tecnica 200 AVS ski boot may only have been produced by one company for one year, but grapnel anchors have been produced by various companies for years. [Good point.](#)

When I realized that I had to reconstruct the entire hierarchy, decided to play a child’s game. I cut paper into small slips and wrote out a vocabulary term on each slip. [This is actually a very useful technique](#) Then I put all the slips in piles based on some inner sense of categories. The categories ended up being – sports (by name), sports equipment, sports techniques, and sports vacation places.

With this new organization, each relationship was genus-species [and](#) or (a [relationship can’t be both, just one or the other](#)) part-whole. The next step was to move these piles of paper into an indented list with bushes based on the three, highest level categories: sports (by name), sports equipment, sports techniques, and sports vacations places. In order to make the bushes clear to users, I added terms that illustrated the

hierarchy but were not concepts from our current database of articles. This allowed for future growth.

I was interesting in checking my vocabulary against an existing sports vocabulary, so I examined Library of Congress subject headings, Yahoo subject director for outdoor sports and one other Internet directory, which may have been Google or Infomine. These directories gave me a better idea of terms that are more commonly used, like horse sports for horseback riding sports, and photography of sports for sports photography. LCSH also showed me a way to break down the top term “sports” into subcategories of sports. In LCSH, narrower terms for sports are: Aquatic sports, Endurance sports, Horse sports, and Winter sports among others. I replaced these as the top level categories, but eventually eliminated all but the Horse sports category. I also eliminated the top term categories of Sports equipment and Sports techniques, which changed the appearance of the hierarchy list. The new organization allowed common topics to appear together, such as all Scuba diving related topics appeared near each other, including the Horse sports bush. I could go one step further and eliminate the Sports vacation places and Sports facilities. This way Golf courses would appear with Golfing and Ski resorts would appear with Skiing. This would make the hierarchical **thesaurus display** closer to the alphabetical **thesaurus display**. True – on the other hand, the two displays serve different purposes, so there’s no need for them to be overly similar.

The process of building the alphabetical thesaurus was pretty interesting. I converted the hierarchical list to a totally flat list, which I moved into a .txt file and then imported into Excel and sorted alphabetically. (Do you know that you can do alphabetical sorts in Word? Create a flat list like you did, then highlight the terms you want to sort, and go to Sort under Table. Even though they’re not in a table, it will sort them. I just discovered this one day.) Then I moved that

alpha list back to a Word file and typed in the Broader and Narrower Term relationships. At this point I was able to think about the Related Terms and consulted the Types of Associative Relationships that explained the nature of related terms. [Good. This procedure makes a lot of sense.](#)

Our group had to compromise over those descriptors that could be considered redundant. Lancaster explains in his *Vocabulary Control for Information Retrieval* that terms used in thesauri need to be unique. “Needless redundancy in specialized thesauri should be avoided. In a thesaurus on library science, there would be little point in beginning many of the terms with the word library” (pp. 53-54). Many of the articles in our group dealt with some of the same subject matter. For example, both “Training your barrel horse the Tyson way” and “The flight response” talk about horses, while “Navigation know-how” and “Anchoring” deal with sailing. The twenty-two sports articles were initially divided between our group’s three members, and each member developed the initial descriptors [for](#) their portion of the articles. Some of the descriptors developed by one person were the same or very similar to some of the terms used by someone else, and so these redundant descriptors had to be consolidated.

One enjoyable aspect of this exercise was being able to see how different people could view the same article from so many different points of view. “Forget the sharks, watch out for turtles,” for example, deals with themes ranging from recreational diving to underwater photography to marine ecology to animal behavior. Whichever theme is emphasized greatly influences the kinds of descriptors developed and the overall effectiveness of the thesaurus. Having to understand these multiple points of view was thought provoking. [good. ultimately, it’s healthy for the end product, too.](#)

There were several minor difficulties that challenged our group, including some of the specific sports terminology relating to subjects like horseback riding. Confusion regarding definitions made it harder to arrange the terms hierarchically because it was not clear if one term was broader than or narrower than another term. [Indexers learn quickly to verify what your sense of a term might mean with a Google search or reference books.](#) In this regard, it may have been more enjoyable to design a thesaurus for articles about topics our members were more familiar with or topics we found more enjoyable. [Yes, I think it would. But indexers and vocabulary have to be able to index in a variety of subject areas.](#)

The thesaurus is to a large extent structurally correct. [A few incorrect relationships and poorly worded terms are present. The discussion section shows some confusion about partitive and instance relationships.](#)

[I liked your jpeg. Visualization really helps sometimes.](#)