Knowledge-Based Spatial Reasoning for Scene Generation from Text Descriptions

Dan Tappan

College of Engineering, Idaho State University

Introduction

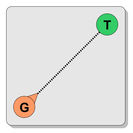
This system translates basic English descriptions of objects in a simplistic zoo environment into plausible, three-dimensional, interactive visualizations of their positions, orientations, and dimensions. It combines a semantic network and contextually sensitive knowledge base as representations for explicit and implicit spatial knowledge, respectively. Its linguistic aspects address underspecification, vagueness, uncertainty, and context with respect to intrinsic, extrinsic, and deictic frames of spatial reference. The underlying, commonsense reasoning formalism is probability-based geometric fields that are solved through constraint satisfaction.

The architecture serves as an extensible test-and-evaluation framework for a multitude of linguistic and artificial-intelligence investigations.

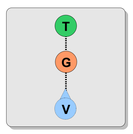
Frames of Spatial Reference

English uses three frames of reference to describe the world [2,3]. Properly interpreting a description depends on selecting the correct one:

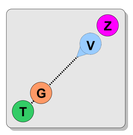
- intrinsic / object-centered
  the tree is in front of the giraffe



- deictic / viewer-centered
  the giraffe is in front of the tree



- extrinsic / environment-centered
  the giraffe is in front of the tree as seen from the zebra

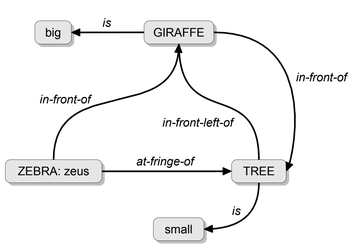


Read

The input comes in a packaged form called a *vignette*. It contains the properly formatted English text, as well as the configuration for any experiments to run.

```
The scene contains a tree, a zebra named Zeus,
  and a giraffe.
Zeus is in front of the giraffe.
Zeus is at the fringe of the tree.
The giraffe is in front of the tree.
The giraffe is big.
The tree is in front and left of the giraffe.
The tree is small.
```

The text is transformed into a semantic network, which is more convenient to manipulate:

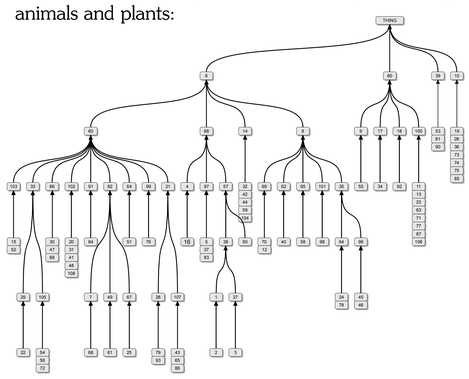


Interpret

Four knowledge-based linguistic issues must be addressed for any interpretation [5]:

- *underspecification* (lack of complete details in a description) requires commonsense world knowledge
- *vagueness* (imprecise nature of details) requires knowledge defining a range of plausible interpretations
- *uncertainty* (lack of commitment to particular interpretations) requires knowledge of preferences
- *context* (different interpretations of objects in certain combinations) requires knowledge to identify the patterns

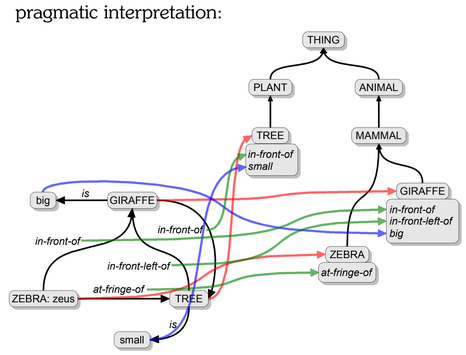
The knowledge base with this information is an inheritance hierarchy of 108 concepts, mostly animals and plants:



#	Concept	#	Concept	#	Concept
1	alligator	37	garter-snake	73	park-bench
2	american-alligator	38	gator	74	pen
3	american-crocodile	39	geographic-thing	75	pickup-truck
4	amphibian	40	giant-manta	76	pig
5	anaconda	41	giraffe	77	pine-tree
6	animal	42	golden-eagle	78	pink-salmon
7	ape	43	gray-wolf	79	pit bull
8	aquatic-animal	44	great-egret	80	plant
9	aquatic-plant	45	great-white-shark	81	pond
10	artificial-thing	46	hammerhead-shark	82	primate
11	aspen-tree	47	hippo	83	python
12	atlantic-octopus	48	horse	84	rabbit
13	birch-tree	49	human	85	raft
14	bird	50	iguana	86	red-wolf
15	blue-whale	51	kangaroo	87	redwood-tree
16	bullfrog	52	killer-whale	88	reptile
17	bush	53	lake	89	rhino
18	cactus	54	leopard	90	river
19	cage	55	lily-pad	91	rodent
20	camel	56	lion	92	saguaro
21	canine	57	lizard	93	saint-bernard
22	cat	58	loch-ness-monster	94	salmon
23	cherry-tree	59	mallard-duck	95	sea-monster
24	coho-salmon	60	mammal	96	shark
25	colobus-monkey	61	man	97	snake
26	corral	62	manta	98	snapping-turtle
27	crocodile	63	maple-tree	99	swine
28	dog	64	marsupial	100	tree
29	domestic-cat	65	mexican-wolf	101	turtle
30	elephant	66	mondopod	102	ungulate
31	elk	67	monkey	103	whale
32	emperor-penguin	68	mountain-gorilla	104	white-pelican
33	feline	69	octopus	105	wild-cat
34	fern	70	pacific-octopus	106	willow-tree
35	fish	71	palm-tree	107	wolf
36	fountain	72	panther	108	zebra

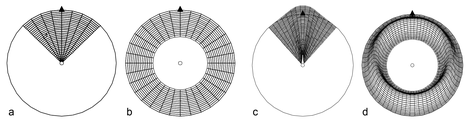
Each concept contains relevant knowledge for its spatial interpretation, such as general spatial characteristics (does it have a face?), prototypical dimensions, and contextually appropriate geometric constraints.

Linking the objects in the semantic network to their concepts in the knowledge base produces a rich, tangled network of meaning for semantic and pragmatic interpretation:

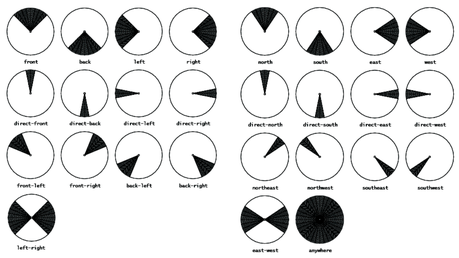


Reason

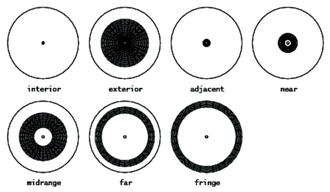
Geometric constraints are defined as circular fields of wedges and rings that specify where other objects may appear with respect to the object in their center. Each field has two complementary parts that specify the legal positions (a-b) and the preferred positions (c-d) [4,7,1,6]:



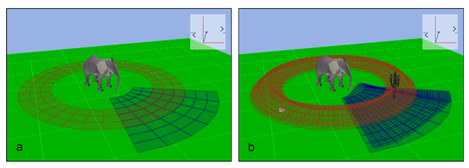
Position and orientation fields use wedges:



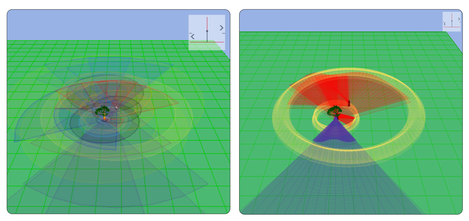
Distance fields use rings:



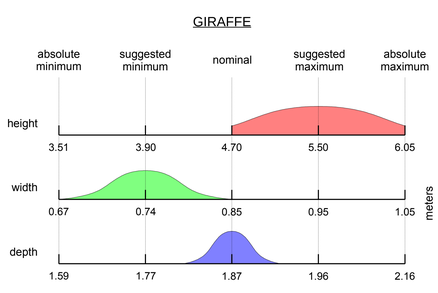
Field intersections specify legal and preferred locations; e.g., *the cactus is in front of* (a, blue wedge) *and near* (a, red ring) *the elephant*. In particular, the higher locations in (b) are more likely.



Extended examples further illustrate the process:

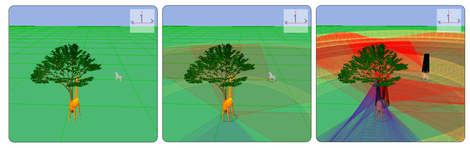


Field diameters are contextually determined by the prototypical dimensions of concepts:



Depict, Infer, and Augment

Running nondeterministic constraint satisfaction over the probability distributions of the fields produces any number of valid interpretations, which can be depicted in various forms:



Finally, querying the positions in every object pairing with respect to their fields generates numerous implicit commonsense spatial relationships from the description [5]:

tree southwest-of world-center
tree far-from world-center
tree local-in-front-of giraffe
tree local-in-front-left-of giraffe
tree global-in-back-of giraffe
tree global-directly-in-back-of giraffe
tree north-of giraffe
tree directly-north-of giraffe
tree outside giraffe
tree near giraffe
tree has-more-height giraffe
tree has-less-width giraffe
tree has-less-depth giraffe
tree local-in-front-of zeus
tree local-directly-in-front-of zeus
tree global-left-of zeus
tree west-of zeus
tree outside zeus
tree near zeus
tree has-more-height zeus
tree has-less-width zeus
tree has-less-depth zeus

giraffe south-of tree
giraffe directly-south-of tree
giraffe at-fringe-of tree
giraffe facing tree
giraffe has-more-width tree
giraffe has-more-depth tree
giraffe has-less-height tree
giraffe south-of world-center
giraffe far-from world-center
giraffe local-left-of zeus
giraffe local-in-front-left-of zeus
giraffe global-in-front-left-of zeus
giraffe southwest-of zeus
giraffe outside zeus
giraffe midrange-from zeus
giraffe facing zeus
giraffe directly-facing zeus
giraffe has-more-height zeus
giraffe has-more-width zeus
giraffe has-more-depth zeus

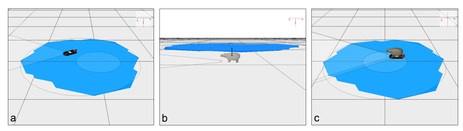
zeus east-of tree
zeus at-fringe-of tree
zeus facing tree
zeus directly-facing tree
zeus has-more-width tree
zeus has-more-depth tree
zeus has-less-height tree
zeus south-of world-center
zeus far-from world-center
zeus local-in-front-of giraffe
zeus local-directly-in-front-of giraffe
zeus global-in-back-right-of giraffe
zeus northeast-of giraffe
zeus outside giraffe
zeus near giraffe
zeus has-less-height giraffe
zeus has-less-width giraffe
zeus has-less-depth giraffe

world-center global-in-back-of giraffe
world-center north-of giraffe
world-center at-fringe-of giraffe
world-center global-in-back-of zeus
world-center north-of zeus
world-center at-fringe-of zeus

Examples

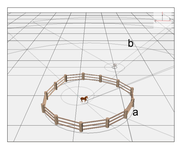
Some effects of context:

a) *The raft is in the lake*
b) *The hippo is in the lake*
c) *The hippo is in the raft, and the raft is in the lake*



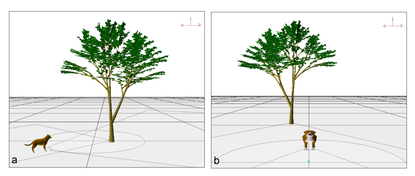
Inclusion and exclusion for circular boundaries:

a) *The horse is inside the corral*
b) *The zebra is outside the corral*



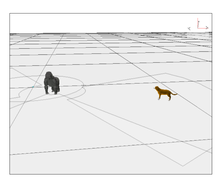
Intrinsic and deictic interpretations:

a) *The tree is in front of the dog*
b) *The dog is in front of the tree*



Side means either left or right:

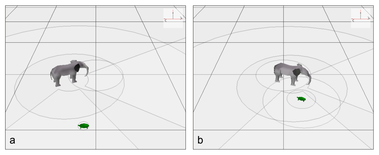
The dog is to the side of the gorilla



Examples

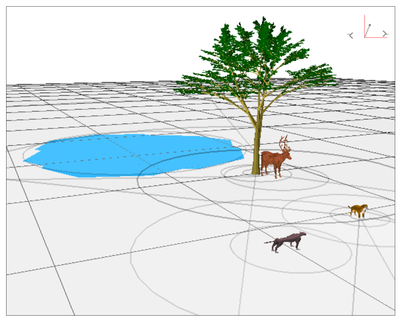
Object dimensions affect context:

a) *The turtle is near the elephant*
b) *The elephant is near the turtle*



Multiple, simultaneous constraints:

The dog is south of the tree and near the panther; the panther is to the right of the dog; and the elk is next to the maple tree and near and facing away from the small pond



Conclusion

This test-and-evaluation framework has proved itself effective in interpreting a wide range of real-world concepts. It addresses limited, yet very practical, contextual aspects of language and space. It also generates substantial inferred knowledge about spatial descriptions, which can be useful for other applications [4,5].

References

[1] Gapp, K. 1994. Basic Meanings of Spatial Relations: Computation and Evaluation in 3D Space. In Proc. AAAI-94, Seattle, WA.

[2] Herskovits, A. 1986. *Language and Spatial Cognition: An interdisciplinary Study of the Prepositions in English*. Cambridge: Cambridge University Press.

[3] Olivier, P. and Tsujii, J. 1994. A computational view of the cognitive semantics of spatial prepositions. In Proc. ACL-94, Las Cruces, NM.

[4] Tappan, D. 2004. Knowledge-Based Spatial Reasoning for Automated Scene Generation from Text Descriptions. Ph.D. diss., New Mexico State University.

[5] Tappan, D. 2008. Augmentation of Explicit Spatial Configuration by Knowledge-Based Inference on Geometric Fields. In Proc. KGCM-08, Orlando, FL.

[6] Tversky, B. 2000. Levels and structure of spatial knowledge. In *Cognitive Mapping: Past, present and future*, Kitchin, R. and Freundshuh, S., eds. London and New York: Routledge.

[7] Yamada, A. 1993. Studies on Spatial Description Understanding Based on Geometric Constraints Satisfaction. Ph.D. diss., Univ. of Kyoto.