Modelling flower colour: Several experiments

Shenghui Wang and David E. Rydeheard*

1 Colour

In order to study the semantics of colour expressions as they are employed in botantical texts, we have analysed a range of flower colours. In this brief note, we describe this and some of the results. Before doing so, we mention a few relevant points about colour in nature.

Colour in materials is usually due to pigments. Pigments that provide colour in plants (and more generally) are well studied and their chemical structures and absorption spectra are known. In plants, there are several families of pigments, those most important for flower colour being the anthocyanins, the flavonols, the flavones and the betalains [37] (see Figure 1 for an example of an anthocyanin molecular structure[1]). Thus any colour we are describing is due to the presence of one or more of these pigments and their chemical environment. This information fixes the colour and so provides an alternative method of defining plant colours – one that is unwieldy, generally inaccessible and far from our own perception!

What we mean by a colour expression is our own perceived response to the 'actual colour' (i.e. the wavelengths of light). Thus one important factor in the semantics of colour is the physiology of human perception of colour. This has been extensively studied (see, for example, [5]). We have three types of colour receptors with peak responses (in normal illumination) in yellow-green, green and blue wavelengths. The consequence of this is that we are more sensitive to differences in shades of green than in, say, red. Different individuals respond differently to colours, at the extreme there are the colour blindnesses but more generally there can be disagreements as to whether, say, an orange is more yellow or more red, or a turquoise more green or more blue. Clearly, in describing colours in plants or other materials, the physiology of human colour perception and the variation in our responses have an important role.

Through the eyes of different creatures, flowers look considerably different. For example, for insect pollinators, there are sometimes patterns called 'pollination guides' in the near UV, invisible to us. Birds often have four colour receptors and most mammals have, unlike ourselves, only two. It is possible to adapt the way we model colours below for other creatures. For example, the three parameter colour models we describe for human perception would be replaced with a four parameter model for birds and a two parameter model for most mammals. Whilst this has considerable biological interest, it is some way from the linguistics of colour!

*School of Computer Science, University of Manchester, Oxford Road, Manchester M13 9PL, U.K. emails: shenghui.wang@cs.man.ac.uk, david.rydeheard@manchester.ac.uk.

[1]Figure transcribed from Wikipedia. Wikipedia copyright applies, see http://en.wikipedia.org.

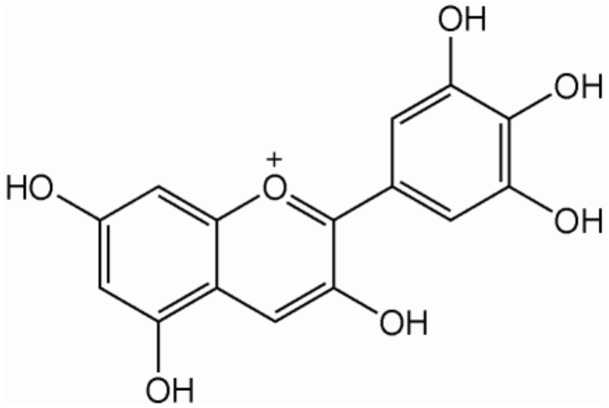

Figure 1: The 2-D structure of the plant pigment Delphinidin present in Delphinium and Viola flowers.

As mentioned before, the use of colour terms in various spoken languages has attracted a good deal of attention. The data about this usage appear to contribute to our understanding of the relationship between 'concept' and 'language', see [8] for further discussion of this aspect of linguistics.

Finally, botanical texts usually use normal everyday terminology for colours, without any attempt to define the terms. In this respect, colour is very different from shape, where there is a specialist and defined terminology. (One may speculate on the reasons for this difference in treatment. Amongst the factors are the capabilities of early printing technology, and the differing roles of shape and colour in plant descriptions, the variability of colour in nature, and the lack of a good everyday nomenclature for shape.) However, alternatives are available. For example, there is the excellent Royal Horticultural Society's (RHS) Colour Chart [36] listing 884 different colours. This is rarely used in texts for wild species, but has been adopted for describing the colour of cultivars. Colour charts for more general purposes are also available, see, for example, the 'Pantone' colours (www.pantone.com). There is also a proposal for a standard 'colour naming scheme', giving definitions of colour terms, see [3], [4], [10] for details, and the website http://tx4.us/moacolor.htm for many colour charts and schemes.

1.1 Colour models

Let us now turn to modelling colour to provide a semantics for colour expressions. Amongst the possible models are:

1. Spectra of visible light.
2. Standard three-parameter colour models. These include the Red-Green-Blue (RGB) model, the Hue-Saturation-Lightness (HSL) model, the artists' model (Red-Yellow-Blue), and the Hue-Chroma-Lightness (HCL) model amongst others (see [16] for an introduction to some of these models).

We deal here with (2) above, leaving light spectra to possible further study.

The Red-Green-Blue (RGB) model is a simple additive model. The colours Red, Green and Blue are first fixed (the criteria for this depend upon application). Any colour is then a combination of these three colours. Usually, each of R, G and B has a fixed range, for example 0-100, or for computer applications using 8-digit binary, 0-255. Thus, pure blue is [0,0,255], pure black is [0,0,0] and pure white is [255,255,255]. Turquoise is something like [72,209,204]. Note that this 'primary colour' model is not that of mixing paints which is the so-called 'artists' model' (Red-Yellow-Blue).

A quite different model is the Hue-Saturation-Lightness (HSL) model. This classifies colours not by contributory colour components, but by three aspects which are meant to represent better how we analyse and match colours. The three-dimensional cartesian space of RGB is replaced by the polar co-ordinates (s, h) with the hue h representing the angle around the 'colour-wheel', and the saturation s the distance from the centre of the wheel. The lightness is orthogonal to this and makes the colour-wheel into a double cone with apices being black (lightness is 0) and white (lightness is 1).

The appropriate metric for the HSL space is the polar expression for Euclidean distance:

$$\mu((h,s,l),(h',s',l')) = \sqrt{(l-l')^2 + (s^2 + s'^2 - 2ss'\cos(h-h'))}.$$

The second component in this expression is the distance between the two vectors (s, h) and (s', h') in polar co-ordinates, and the first component is the Euclidean addition of the lightness. Although this is an adaptation of Euclidean distance in HSL, it translates into a non-Euclidean metric in RGB.

The geodesics of this metric are indeed straight lines (not in (h, s, l) parameters, of course, but in the combination of polar and cartesian that these represent). Parameterised by t, the geodesics are:

$$\begin{aligned} s\cos(h) &= \alpha t + \alpha_0 \\ s\sin(h) &= \beta t + \beta_0 \\ l &= \gamma t + \gamma_0 \end{aligned}$$

where $\alpha, \alpha_0, \beta, \beta_0, \gamma, \gamma_0$ are constants. These geodesics when depicted do appear to be in accord with our expectations of colours in descriptions of colour ranges.

Whilst the HSL model and its metric are closer to human colour perception than RGB, variant models have been proposed which are claimed to be even closer to our colour perception with metrics more exactly matching our perceived proximity of colours. For example, Sarifuddin and Missaoui [38] propose a Hue-Chroma-Luminance (HCL) model. The HCL model is similar to the HSL model in that hue is represented as an angle on the 'colour-wheel' and the second component, the chroma, is the distance from the origin in this disk. The luminance measures the way we respond to colour intensity, and provides an orthogonal axis making the disk into a cone. Though the parameters provide a similar shape of model to HSL, the way each component is defined in terms of RGB values is different. Their proposed metric is a weighted version of the HSL metric:

$$\mu((h,c,l),(h',c',l')) = \sqrt{\lambda_1(l-l')^2 + \lambda_2(c^2 + c'^2 - 2cc'\cos(h-h'))}.$$

The two λs are factors which are chosen to make this distance reflect perceptual proximity. Suggested values are:

$$\lambda_1 = 1.4456 \qquad \lambda_2 = |h-h'| + 0.16$$

where h is expressed in degrees.

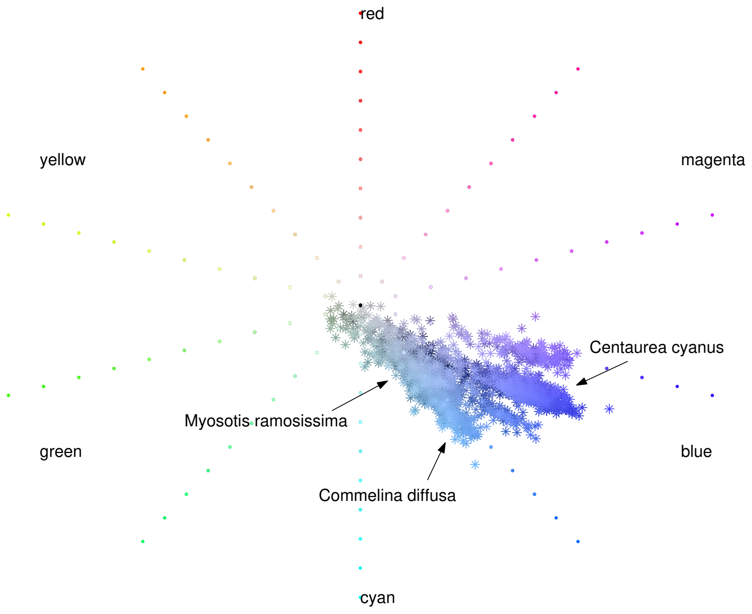

Figure 2: HSL plot of pixel colours of various blue flowers.

All the three-parameter models are equivalent in the sense that there are bijective conversions between them. However, some are closer to our visual perception and the way we separate the components of a colour. Such models are more relevant to a description of semantics. Moreover, standard metrics in some models do not reflect the semantics of colour, for example, the Euclidean metric in the the RGB model does not correctly capture perceived colour differences.

1.2 Semantics of colour

We now explore the possibility of using these models to give a semantics of colour expressions as they occur in botany. As an example, we use the HSL model and its metric – it being a standard model reflecting, to some extent, our visual perception. Colour phrases are interpreted in this model using standard interpretations for basic colour terms *red*, *blue* etc. and for colour modifiers *pale*, *deep* etc. (see, for example, [3] and [4]) and, for combined expressions, the general metric semantics of continuous quantities.

To compare the semantics with actual colours in nature, we adopt a simple strategy. We use digital images of flower colours and analyse them to find the position of the colour components of pixels in the colour model. As mentioned previously, there is a good deal of processing and transformation of colour in these images from the original reflected light to the image processed for colour information, and indeed the camera itself is adapted to our own colour responses.

Let's begin with simple colour words as they are used in botany. In Figure 2, we depict the pixel colours of three blue flowers in the Hue-Saturation-Lightness model. The image is a projection of the HSL double cone onto its 'base', i.e. depicts only the variation of Hue with Saturation. The saturation varies from zero at the centre to maximum at the perimeter.

The flowers analysed are (1) *Centaurea cyanus* (Cornflower) whose major pigment is Protocyanin [42] and whose flower colour is variously described as:

> blue
> violet-blue or dark blue, rarely white or purple
> violet-blue, outer dark blue, white or purple
> usually blue, sometimes pink or white
> blue,

and (2) *Myosotis ramosissima* (Early Forget-me-not), variously described as:

> mid-blue
> bright blue
> blue (rarely white),

and finally (3) *Commelina diffusa* (Climbing Dayflower), which is described as:

> upper petals blue
> bright blue, or rarely white
> medium blue
> blue.

What is clear from the colour plot is that the colours of the flowers occupy different regions of the HSL space, all near blue. Some of the authors quoted have attempted to describe more precisely these regions, either by saying which colour areas to move towards to modify pure blue (e.g. *violet-blue*) or with modifiers such as *mid*, *medium*, *bright* or *dark*. Notice also that the degree of saturation is fairly low, and varies between species, with Cornflower having a greater saturation (maximizing at approx. 60%) – a fact which is less apparent from the botanical descriptions.

Let us now turn to descriptions of colour variation. As an example of a plant whose flowers exhibit a range of colours, consider *Pulmonaria longifolia* (Narrow-leaved Lungwort). This is described as:

> more vivid blue [than *P. officinalis*]
> red at first, soon turning to violet or violet-blue
> violet to violet-blue
> blue to violet when open.

Notice that the color range is here mixed with flower maturity in (a product metric). In Figure 3, we depict the flower colour in HSL for flowers of differing ages. The first plot is the the projection of the double cone onto its base, i.e. the variation of Hue with Saturation, displayed as in the previous plot. The second plot is the projection of the HSL double cone on the red-cyan axis, with red on the right. As expected, the plots display a much greater range of colours than is evident in the blue flowers previously discussed. The first plot shows that the hues are indeed in a range from fairly pure blue through to red (on the blue side of reds) and the second plot shows that the red pixels (on the right) are predominantly in the darkish red area (the bottom centre is the lower apex of the cone and is black, the top centre being white), whilst the blue pixels (on the left) range more widely, especially in the lighter blues where they extend out to a saturation of nearly 60%.

Several things are not clear from these plots. For example, how frequent are the reds and the blues, and is the frequency fairly even between these or does it peak? In Figure 4,

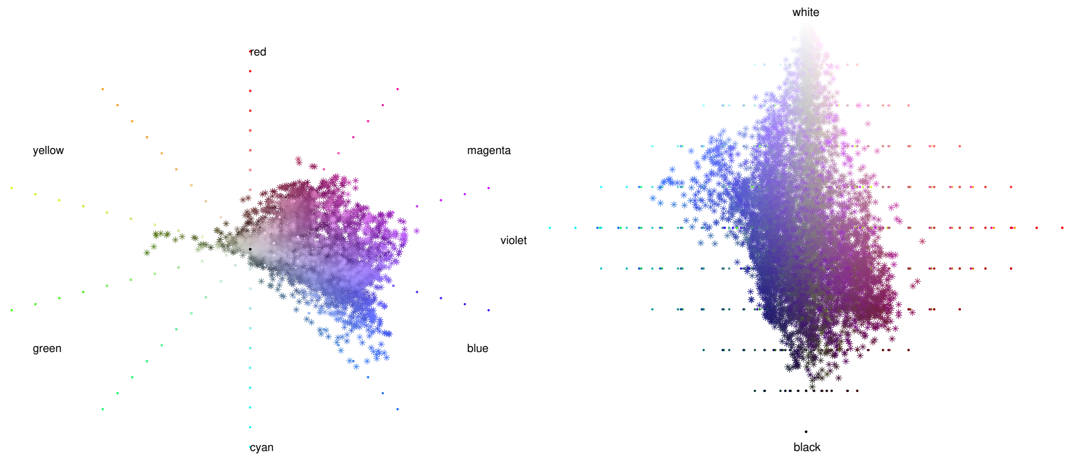

Figure 3: HSL plots of the flower colour pixels for *Pulmonaria longifolia*.

we plot the frequencies of pixels against the hue, with the x-axis denoting the percentage distance around the colour wheel. We notice that the distribution is bimodal, with two peaks, one in the blue (the first narrower peak at c68), and one in purplish red area (the second broader peak at c90) and a marked dearth of pixels of intermediate hue. This corresponds to the observation that these flowers are initially purplish red in bud, but then rapidly progress through intermediate colours to stabilise around the blue peak.

Let us look at another colour range in flowers. In Figure 5, we plot the same HSL configurations for some digital images of flowers of *Rhododendron elegantulum*. This flower colour is variously described as:

> pale pink
> pale pinkish purple or white flushed pink with crimson spots
> pink to white strongly flushed pink with reddish spots.

What do we see from the plots? There is a fairly large region of HSL space occupied by the pixels, though a narrowish band of hues on the blue-purple side of red. Most of the spread is along the saturation axis and, in the second plot, along the lightness: with black at the bottom and white at the top, we see that lighter shades tend to predominate with some white present.

It is clear that these results reflect a much broader colour range than is evident from the semantics of colour descriptions. There are several reasons for this. Firstly, there is the paucity of our standard vocabulary for colours and colour ranges, and our meagre use of precision in everyday descriptions of colour, and the concentration of this description on the hue component. Thus botanists are unlikely to attempt to capture the real variation present in these plots, even if they chose to do so. In this, we are in a similar situation to leaf shapes, as we are with the multiplicity of different parallel colour descriptions, and the possibility of integration using the semantics. Moreover, it is not clear how much the ranges in the plots are artifices of the colour processing in digital images. Of course, this doesn't preclude our trying to describe these ranges, but they may not exactly match what a botanist is seeing. To reach a conclusion here, spectrographic analysis and/or

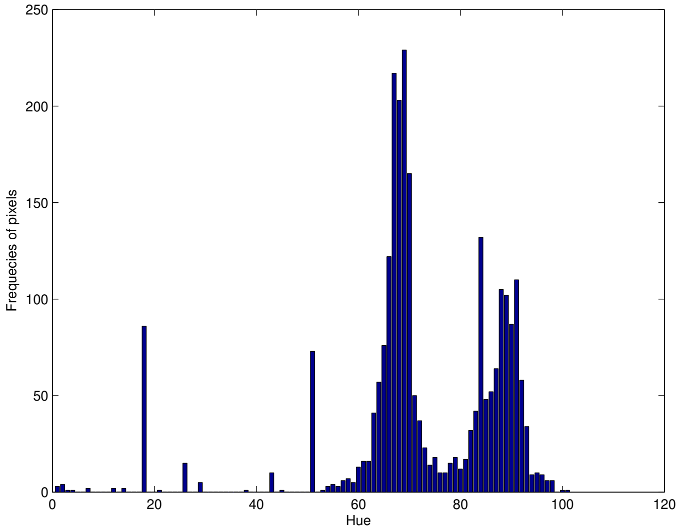

Figure 4: Plot of frequency against hue for the colour pixels of *Pulmonaria longifolia*.

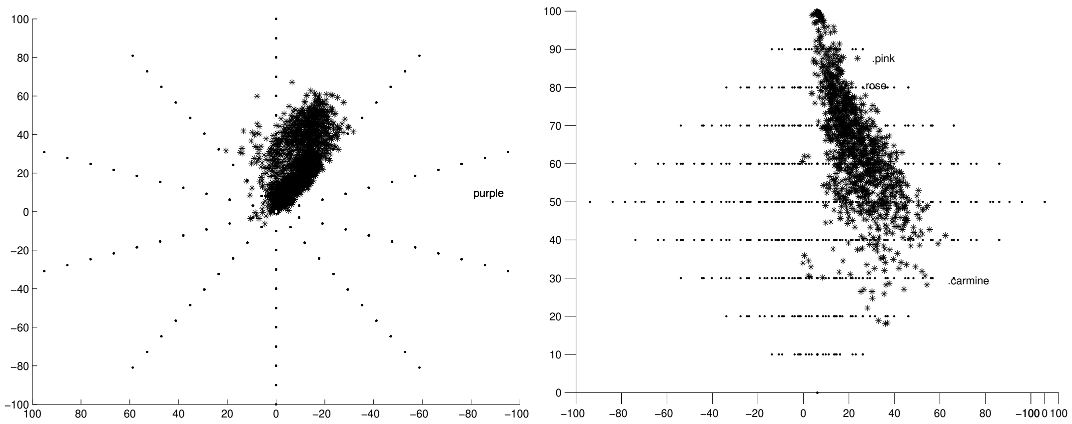

Figure 5: HSL plots of the flower colour pixels for *Rhododendron elegantulum*.

knowledge of the combinations of pigments present is necessary. However, even this has to be modified by the profile of the colour receptors of our eyes and the way we process these signals.

This is clearly only the beginning of an analysis of colour expressions in botany. The techniques described above suggest a series of further experiments in this practical approach to semantics.

References

[1] J.W. de Bakker and E.P. de Vink. *Control Flow Semantics.* Foundations of Computing Series, MIT Press. Cambridge, MA. 1996.

[2] S. Bechhofer, F. van Harmelen, J. Henderson, I. Horrocks, D.L. McGuiness, P.F. Patel-Schneider, and eds., L.A.S. (2004) OWL Web Ontology Language Reference. http://www.w3.org/TR/owl-ref.

[3] T. Berk, L. Brownston, and A. Kaufman. A new color-naming scheme for graphics languages. *IEEE CG&A.* Vol 2. No. 3. pp 37-44. IEEE 1982.

[4] T. Berk, L. Brownston, and A. Kaufman. A human factors study of color notation systems for computer graphics. *Comm. ACM.* Vol 25. No. 8. pp 547-550. ACM, August 1982.

[5] C.M. Bloomer. *Principles of Visual Perception.* New York: Van Nostrand. 1976.

[6] F.L. Bookstein. *Morphometric tools for landmark data: Geometry and biology.* Cambridge: Cambridge University Press. 1991.

[7] K. B orner. Extracting and Visualizing Semantic Structures in Retrieval Results for Browsing. *International Conference on Digital Libraries.* Proc. 5th ACM Conference on Digital Libraries. San Antonio, Texas. pp 234-235. 2000.

[8] Alex Byrne and David Hilbert. A Bibliography of Color and Philosophy. *Readings on Color, Vol. 1: The Philosophy of Color.* MIT Press. 1997 (See also: http://web.mit.edu/philos/www/color-biblio.html)

[9] A.R. Clapham, T.G. Tutin and E.F. Warburg. *Excursion Flora of the British Isles.* Cambridge University Press, 1973.

[10] D. Conway. An experimental comparison of three natural language colour naming models. *Proc. East-West International Conference on Human-Computer Interaction.* St. Petersburg, Russia. pp 328-339. 1992.

[11] E.T. Copson. *Metric Spaces.* Cambridge Tracts in Mathematics 57, Cambridge University Press. 1968.

[12] T.A. Dickinson, W.H. Parker and R.E. Strauss. Another approach to leaf shape comparisons. *Taxon* 36(1): 1-20. 1987.

[13] Mike Dowman. A Bayesian Approach To Colour Term Semantics. Lingscene, Volume 1. 2001.

[14] Shimon Edelman. Representation, Similarity, and the Chorus of Prototypes. *Minds and Machines* 5:45-68. 1995.

[15] Max J. Egenhofer and A. Rashid B.M. Shariff. Metric Details for Natural-Language Spatial Relations. *ACM Trans. Inf. Syst.* 16(4) 295-321. 1998.

[16] James D. Foley, Andries van Dam, Steven K. Feiner and John F. Hughes. *Computer Graphics: Principles and Practice.* Addison-Wesley, 1996.

[17] Peter G ardenfors. *Conceptual Spaces.* MIT Press. Cambridge, MA. 2000.

[18] Johan Gielis. A Generic Transformation that Unifies a Wide Range of Natural and Abstract Shapes. *American Journal of Botany*, 90(3), 333-338. 2003.

[19] Johan Gielis and Tom Gerats. A botanical perspective on modeling plants and plant shapes in computer graphics. Radboud University, Nijmegen, Netherlands. http://www.pg.science.ru.nl/en/plantmodeling.html.

[20] J.R. Giles. *Introduction to the Analysis of Metric Spaces.* Australian Mathematical Society, Lecture Series 3. Cambridge University Press.1987.

[21] T. Hofmann. Probabilistic Latent Semantic Analysis. *Uncertainty in Artificial Intelligence, UAI'99.* Stockholm. 1999.

[22] Gabriel Lamé. Examen des différentes méthodes employées pour résoudre les problèmes de geometrie. Paris: Courcier, 1818.

[23] Ales Jaklic, Ales Leonardis and Franc Solina. *Segmentation and Recovery of Superquadrics* Computational imaging and vision, Vol. 20, Kluwer, Dordrecth, 2000.

[24] J.M. Lammens. *A computational model of color perception and color naming.* PhD Thesis. State University of New York at Buffalo. 1994.

[25] T.K. Landauer, P.W. Foltz and D. Laham. Introduction to Latent Semantic Analysis. *Discourse Processes* Vol 25. pp 259-284. 1998. See also the website: http://lsa.colorado.edu.

[26] L.F. Marcus, M. Corti, A. Loy, G.J.P. Naylor, and D.E. Slice, Eds. *Advances in Morphometrics.* NATO Advanced Science Institute Series A: Life Sciences. Volume 284. New York, Plenum Press.

[27] S. Marchand-Maillet et al. Viper's CBIRS webpage. Online address: http://viper.unige.ch/other_systems

[28] Susannah J. Lydon, Mary McGee Wood, Robert Huxley and David Sutton. Data patterns in multiple botanical descriptions: implications for automated processing of legacy data. *Systematics and Biodiversity* 1(2) 151-157. 2003.

[29] Jimmy R. Massey and James C. Murphy. Vascular Plant Systematics: Categorized Glossary. 1998. Online address: http://www.ibiblio.org/botnet/glossary

[30] M. Ossian, T. Tuyelaars and L. Van Gool. Fitting superellipses to incomplete contours. *Conference on Computer Vision and Pattern Recognition Workshop.* (CVPRW'04) Vol. 4. 49-57. 2004.

[31] J.Z. Pan. *Description Logics: Reasoning Support for the Semantic Web.* PhD Thesis, School of Computer Science, The University of Manchester.

[32] P. Pantel and D. Lin. Discovering Word Senses from Text. *Conference on Knowledge Discovery In Data.* Proc. 8th ACM SIGKDD Conference. pp 613-619. 2002.

[33] J.F. Quesada, W. Kintsch and E. Gomez. A theory of Complex Problem Solving using Latent Semantic Analysis. In W.D. Gray and C.D. Schunn (Eds) Proceedings of the *24th Annual Conference of the Cognitive Science Society.* Fairfax VA, Lawrence Erlbaum Assocs, Mahwah, NJ. pp 750-755. 2002.

[34] F.J. Rohlf and L.F. Marcus. A revolution in morphometrics. *Trends in Ecol. Evol.* 8: 129-132.

[35] C.J. van Rijsbergen. *Information Retrieval.* Second Edition. Butterworth. 1979.

[36] RHS Colour Chart. Royal Horticultural Society Publications. See http://www.rhs.org.uk

[37] Frank B. Salisbury and Cleon W. Ross. *Plant Physiology.* Wadsworth Publishing Co. 1992.

[38] M. Sarifuddin and R. Missaoui. A New Perceptually Uniform Colour Space and Associated Similarity Measure for Content-Based Image and Video Retrieval. Multimedia Information Retrieval Workshop. Salvador, Brazil. 2005.

[39] Schwering, Angela. Hybrid Model for Semantic Similarity Measurement. University of M unster. http://ifgi.uni-muenster.de/ eidueidu.

[40] Schwering, A. and Raubal, M. Spatial Relationships for Semantic Similarity Measurement. 2nd International Workshop on Conceptual Modeling for Geographic Information Systems. Klagenfurt, Austria. 2005

[41] E. Sharon and D. Mumford. 2D-Shape Analysis using Conformal Mapping. CVPR (2) 2004: 350-357.

[42] Masaaki Shiono, Naohiro Matsugaki and Kosaku Takeda. Phytochemistry: Structure of the blue cornflower pigment. *Nature* 436, 791 (11 August 2005).

[43] Clive Stace. *New Flora of the British Isles.* Cambridge University Press. 1997.

[44] D'Arcy W. Thompson. *On Growth and Form.* Cambridge University Press, London. 1917.

[45] Shenghui Wang and Jeff Z. Pan. Ontology-based Representation and Query of Colour Descriptions from Botanical Documents. in *Proc. of the 4th International Conference on Ontologies, DataBases, and Applications of Semantics (ODBASE-2005).* 2005.

[46] William T. Stearn *Botanical Latin.* Timber Press, 4th edition, 2004.

[47] Wood, M., Lydon, S., Tablan, V., Maynard, D. and Cunningham, H. Populating a database from parallel texts using ontology-based information extraction. In Meziane, F. and Métais, E., (editors) *Proceedings of Natural Language Processing and Information Systems, 9th International Conference on Applications of Natural Languages to Information Systems.* 254-264. Springer 2004.

[48] Wood, M., and Wang, S. Motivation for "ontology" in parallel-text information extraction. In *Proceedings of ECAI-2004 Workshop on Ontology Learning and Population. (ECAI-OLP).* Poster. Valencia, Spain. 2004.

[49] Wood, M.M., Lydon, S.J., Tablan, V., Maynard, D., and Cunningham, H. Using parallel texts to improve recall in ie. In *Proc. of Recent Advances in Natural Language Processing (RANLP-2003).* 506-512. Borovetz, Bulgaria. 2003.

[50] Yongyue Zhang. Webpage resource on image morphing. http://www.fmrib.ox.ac.uk/ yongyue/morphing.html